

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
20 June 2002 (20.06.2002)

PCT

(10) International Publication Number
WO 02/48370 A2(51) International Patent Classification⁷: C12N 15/12,
C12Q 1/68, C12N 1/21, 5/10, C07K 14/47, 16/18, G01N
33/68, A61K 39/395, 38/17, 31/713LIU, Chenghua [CN/US]; 1125 Ranchero Way #14, San
Jose, CA 95117 (US).

(21) International Application Number: PCT/US01/51341

(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrrell
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(22) International Filing Date: 30 October 2001 (30.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/244,717 31 October 2000 (31.10.2000) US(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/244,717 (CIP)
Filed on 31 October 2000 (31.10.2000)(71) Applicant (for all designated States except US): DI-
ADEXUS, INC. [US/US]; 343 Oyster Point Boulevard,
South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SUN, Yongming
[CN/US]; 551 Shoal Drive, Redwood City, CA 94065
(US). RECIPON, Herve [FR/FR]; 85 Fortuna Avenue,
San Francisco, CA 94115 (US). GHOSH, Malavika, G.
[IN/US]; 1733 Via Lugano, San Jose, CA 95120 (US).(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:

— without international search report and to be republished
upon receipt of that reportFor two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/48370 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic colon cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon tissue, identifying colon tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.

DEX-0523

AB

COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application
5 Serial No. 60/244,717 filed October 31, 2000, which is herein incorporated by reference
in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and
10 polypeptides present in normal and neoplastic colon cells, including fragments, variants
and derivatives of the nucleic acids and polypeptides. The present invention also relates
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of
the polypeptides of the invention. The invention also relates to compositions comprising
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists
15 of the invention and methods for the use of these compositions. These uses include
identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-
cancerous disease states in colon tissue, identifying colon tissue and monitoring and
identifying and/or designing agonists and antagonists of polypeptides of the invention.
The uses also include gene therapy, production of transgenic animals and cells, and
20 production of engineered colon tissue for treatment and research.

BACKGROUND OF THE INVENTION

Colorectal cancer is the second most common cause of cancer death in the United
States and the third most prevalent cancer in both men and women. M. L. Davila & A.
D. Davila, *Screening for Colon and Rectal Cancer*, in Colon and Rectal Cancer 47 (Peter
25 S. Edelstein ed., 2000). Approximately 100,000 patients every year suffer from colon
cancer and approximately half that number die of the disease. Hannah-Ngoc Ha & Bard
C. Cosman, *Treatment of Colon Cancer*, in Colon and Rectal Cancer 157 (Peter S.
Edelstein ed., 2000). Nearly all cases of colorectal cancer arise from adenomatous
polyps, some of which mature into large polyps, undergo abnormal growth and
30 development, and ultimately progress into cancer. Davila & Davila, *supra* at 55-56.
This progression would appear to take at least 10 years in most patients, rendering it a

readily treatable form of cancer if diagnosed early, when the cancer is localized. *Id.* at 56; Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 125 (1998).

Although our understanding of the etiology of colon cancer is undergoing continual refinement, extensive research in this area points to a combination of factors, including age, hereditary and nonhereditary conditions, and environmental/dietary factors. Age is a key risk factor in the development of colorectal cancer, Davila & Davila, *supra* at 48, with men and women over 40 years of age become increasingly susceptible to that cancer, Burdette, *supra* at 126. Incidence rates increase considerably in each subsequent decade of life. Davila et al., *supra* at 48. A number of hereditary and nonhereditary conditions have also been linked to a heightened risk of developing colorectal cancer, including familial adenomatous polyposis (FAP), hereditary nonpolyposis colorectal cancer (Lynch syndrome or HNPCC), a personal and/or family history of colorectal cancer or adenomatous polyps, inflammatory bowel disease, diabetes mellitus, and obesity. *Id.* at 47; Henry T. Lynch & Jane F. Lynch, *Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndromes)*, in Colon and Rectal Cancer 67-68 (Peter S. Edelstein ed., 2000).

In the case of FAP, the tumor suppressor gene APC (adenomatous polyposis coli), located at 5q21, has been either mutationally inactivated or deleted. Alberts et al., Molecular Biology of the Cell 1288 (3d ed. 1994). The APC protein plays a role in a number of functions, including cell adhesion, apoptosis, and repression of the *c-myc* oncogene. N. R. Hall & R. D. Madoff, *Genetics and the Polyp-Cancer Sequence*, Colon and Rectal Cancer 8 (Peter S. Edelstein, ed., 2000). Of those patients with colorectal cancer who have normal APC genes, over 65% have such mutations in the cancer cells but not in other tissues. Alberts et al., *supra* at 1288. In the case of HNPCC, patients manifest abnormalities in the tumor suppressor gene HNPCC, but only about 15% of tumors contain the mutated gene. *Id.* A host of other genes have also been implicated in colorectal cancer, including the *K-ras*, *N-ras*, *H-ras* and *c-myc* oncogenes, and the tumor suppressor genes *DCC* (deleted in colon carcinoma) and *p53*. Hall & Madoff, *supra* at 8-9; Alberts et al., *supra* at 1288.

Environmental/dietary factors associated with an increased risk of colorectal cancer include a high fat diet, intake of high dietary red meat, and sedentary lifestyle. Davila & Davila, *supra* at 47; Reddy, B. S., *Prev. Med.* 16(4): 460-7 (1987).

Conversely, environmental/dietary factors associated with a reduced risk of colorectal cancer include a diet high in fiber, folic acid, calcium, and hormone-replacement therapy in post-menopausal women. Davila & Davila, *supra* at 50-55. The effect of antioxidants in reducing the risk of colon cancer is unclear. *Id.* at 53.

5 Because colon cancer is highly treatable when detected at an early, localized stage, screening should be a part of routine care for all adults starting at age 50, especially those with first-degree relatives with colorectal cancer. One major advantage of colorectal cancer screening over its counterparts in other types of cancer is its ability to not only detect precancerous lesions, but to remove them as well. Davila & Davila,
10 *supra* at 56. The key colorectal cancer screening tests in use today are fecal occult blood test, sigmoidoscopy, colonoscopy, double-contrast barium enema, and the carcinoembryonic antigen (CEA) test. *Id.*; Burdette, *supra* at 125.

 The fecal occult blood test (FOBT) screens for colorectal cancer by detecting the amount of blood in the stool, the premise being that neoplastic tissue, particularly
15 malignant tissue, bleeds more than typical mucosa, with the amount of bleeding increasing with polyp size and cancer stage. Davila & Davila, *supra* at 56-57. While effective at detecting early stage tumors, FOBT is unable to detect adenomatous polyps (premalignant lesions), and, depending on the contents of the fecal sample, is subject to rendering false positives. *Id.* at 56-59. Sigmoidoscopy and colonoscopy, by contrast,
20 allow direct visualization of the bowel, and enable one to detect, biopsy, and remove adenomatous polyps. *Id.* at 59-60, 61. Despite the advantages of these procedures, there are accompanying downsides: sigmoidoscopy, by definition, is limited to the sigmoid colon and below, colonoscopy is a relatively expensive procedure, and both share the risk of possible bowel perforation and hemorrhaging. *Id.* at 59-60. Double-contrast barium
25 enema (DCBE) enables detection of lesions better than FOBT, and almost as well a colonoscopy, but it may be limited in evaluating the winding rectosigmoid region. *Id.* at 60. The CEA blood test, which involves screening the blood for carcinoembryonic antigen, shares the downside of FOBT, in that it is of limited utility in detecting colorectal cancer at an early stage. Burdette, *supra* at 125.

30 Once colon cancer has been diagnosed, treatment decisions are typically made in reference to the stage of cancer progression. A number of techniques are employed to stage the cancer (some of which are also used to screen for colon cancer), including

pathologic examination of resected colon, sigmoidoscopy, colonoscopy, and various imaging techniques. AJCC Cancer Staging Handbook 84 (Irvin D. Fleming et al. eds., 5th ed. 1998); Montgomery, R. C. and Ridge, J.A., *Semin. Surg. Oncol.* 15(3): 143-150 (1998). Moreover, chest films, liver functionality tests, and liver scans are employed to

5 determine the extent of metastasis. Fleming et al. eds., *supra* at 84. While computerized tomography and magnetic resonance imaging are useful in staging colorectal cancer in its later stages, both have unacceptably low staging accuracy for identifying early stages of the disease, due to the difficulty that both methods have in (1) revealing the depth of bowel wall tumor infiltration and (2) diagnosing malignant adenopathy. Thoeni, R. F.,

10 *Radiol. Clin. N. Am.* 35(2): 457-85 (1997). Rather, techniques such as transrectal ultrasound (TRUS) are preferred in this context, although this technique is inaccurate with respect to detecting small lymph nodes that may contain metastases. David Blumberg & Frank G. Opelka, *Neoadjuvant and Adjuvant Therapy for Adenocarcinoma of the Rectum, in Colon and Rectal Cancer* 316 (Peter S. Edelstein ed., 2000).

15 Several classification systems have been devised to stage the extent of colorectal cancer, including the Dukes' system and the more detailed International Union against Cancer-American Joint Committee on Cancer TNM staging system, which is considered by many in the field to be a more useful staging system. Burdette, *supra* at 126-27. The TNM system, which is used for either clinical or pathological staging, is divided into four

20 stages, each of which evaluates the extent of cancer growth with respect to primary tumor (T), regional lymph nodes (N), and distant metastasis (M). Fleming et al. eds., *supra* at 84-85. The system focuses on the extent of tumor invasion into the intestinal wall, invasion of adjacent structures, the number of regional lymph nodes that have been affected, and whether distant metastasis has occurred. *Id.* at 81.

25 Stage 0 is characterized by *in situ* carcinoma (Tis), in which the cancer cells are located inside the glandular basement membrane (intraepithelial) or lamina propria (intramucosal). *Id.* at 84-85; Burdette, *supra* at 127. In this stage, the cancer has not spread to the regional lymph nodes (N0), and there is no distant metastasis (M0). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. In stage I, there is still no spread

30 of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the submucosa (T1) or has progressed further to invade the muscularis propria (T2). Fleming et al. eds., *supra* at 84-85; Burdette, *supra* at 127. Stage II also involves

-5-

no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the subserosa, or the nonperitonealized pericolic or perirectal tissues (T3), or has progressed to invade other organs or structures, and/or has perforated the visceral peritoneum (T4). *Id.* Stage 3 is characterized by any of the T substages, no
5 distant metastasis, and either metastasis in 1 to 3 regional lymph nodes (N1) or metastasis in four or more regional lymph nodes (N2). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. Lastly, stage 4 involves any of the T or N substages, as well as distant metastasis. *Id.*

Currently, pathological staging of colon cancer is preferable over clinical staging
10 as pathological staging provides a more accurate prognosis. Pathological staging typically involves examination of the resected colon section, along with surgical examination of the abdominal cavity. Fleming et al. eds., *supra* at 84. Clinical staging would be a preferred method of staging were it at least as accurate as pathological staging, as it does not depend on the invasive procedures of its counterpart.

15 Turning to the treatment of colorectal cancer, surgical resection results in a cure for roughly 50% of patients. Burdette, *supra* at 125. Irradiation is used both preoperatively and postoperatively in treating colorectal cancer. *Id.* at 125, 132-33. Chemotherapeutic agents, particularly 5-fluorouracil, are also powerful weapons in treating colorectal cancer. *Id.* at 125, 133. Other agents include irinotecan and
20 floxuridine, cisplatin, levamisole, methotrexate, interferon-alpha, and leucovorin. *Id.* at 133. Nonetheless, thirty to forty percent of patients will develop a recurrence of colon cancer following surgical resection. Wayne De Vos, *Follow-up After Treatment of Colon Cancer, Colon and Rectal Cancer* 225 (Peter S. Edelstein ed., 2000), which in many patients is the ultimate cause of death. Accordingly, colon cancer patients must be
25 closely monitored to determine response to therapy and to detect persistent or recurrent disease and metastasis.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of colorectal cancer are of critical importance to the outcome of the patient. Moreover, current procedures,
30 while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that

-6-

would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop colorectal cancer, for diagnosing colorectal cancer, for monitoring the progression of the disease, for staging the colorectal cancer, for determining whether the colorectal cancer has metastasized, and for imaging the colorectal cancer. There is also a need for better treatment of colorectal cancer.

SUMMARY OF THE INVENTION

10 The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon; identify and monitor colon tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered colon tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to colon cells and/or colon tissue. These colon specific nucleic acids (CSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the CSNA is genomic DNA, then the CSNA is a colon specific gene (CSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 148 through 250. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 147. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a CSP, or that selectively hybridize or exhibit substantial sequence similarity to a CSNA, as well as allelic variants of a nucleic acid molecule encoding a CSP, and allelic variants of a CSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a CSP or that comprises a part of a nucleic acid sequence of a CSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a CSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a CSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a CSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a CSNA.

Another object of the invention is to provide methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a CSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a CSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring colon tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered colon tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon. The invention provides methods of using the polypeptides of the invention to identify and/or monitor colon tissue, and to produce engineered colon tissue.

5 The agonists and antagonists of the instant invention may be used to treat colon cancer and non-cancerous disease states in colon and to produce engineered colon tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for
10 comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by
15 those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used
20 in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and
25 Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory
30 Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 10 A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and
- 15 "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 20 The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,
- 25 phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially
- 30 duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are

known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site,

recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by
5 expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as
10 those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other
15 polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages
20 such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See *e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical
25 Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In
30 addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each

nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

5 The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

 The term "percent sequence identity" in the context of nucleic acid sequences
10 refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number
15 of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity
20 of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent
25 sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

 A reference to a nucleic acid sequence encompasses its complement unless
30 otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its

complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"

- 5 interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its
10 complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

- 15 Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at
20 least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

- Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species,
25 length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of
30 hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is

-15-

performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

10 The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

15 In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

25 An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours).
30 Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten

hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify
5 nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents
10 to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook
15 (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in
20 the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon
25 degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula:
$$T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/\text{N}),$$

wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p.
30 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using

mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g., Ausubel (1999), supra; Sambrook (1989), supra, pp. 11.45-11.57.*

5 The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment
10 is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are
15 specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the
20 art.

 The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g., Sambrook (1989), supra.*

25 Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon
30 portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-

-18-

derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one
5 exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic
10 acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing
15 Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that
20 nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the
25 nucleic acid molecule comprises the wild type nucleic acid sequence encoding a CSP or is a CSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high
30 rate of point mutations is obtained along the entire length of the PCR product. See, e.g., Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. See, e.g., Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

5 The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

10 The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. See, e.g., Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

15 The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

20 The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

25 The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815
30 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein

small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g., Delegrave et al., Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in
5 its entirety.

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

10 The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,
15 promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such
20 control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term “control sequences” is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

25 The term “vector,” as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of
30 vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are

introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “recombinant expression vectors” (or simply, “expression vectors”). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, “plasmid” and “vector” may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term “recombinant host cell” (or simply “host cell”), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term “host cell” as used herein.

As used herein, the phrase “open reading frame” and the equivalent acronym “ORF” refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase “ORF-encoded peptide” refers to the predicted or actual translation of an ORF.

As used herein, the phrase “degenerate variant” of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term “polypeptide” encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide

each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a CSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide
5 that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be
10 "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single
15 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,
20 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared
25 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40
30 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

-23-

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a

different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs.

- 5 The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide
- 10 analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

- The term "non-peptide analog" refers to a compound with properties that are
- 15 analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm
- 20 polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: $-\text{CH}_2\text{NH}-$, $-\text{CH}_2\text{S}-$, $-\text{CH}_2-\text{CH}_2-$, $-\text{CH}=\text{CH}-$ (*cis* and *trans*), $-\text{COCH}_2-$, $-\text{CH}(\text{OH})\text{CH}_2-$, and $-\text{CH}_2\text{SO}-$, by methods well-known in the art. Systematic substitution of one or more amino acids of a
- 25 consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may
- 30 add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

-25-

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (e.g., a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure,

Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), *Immunology - A Synthesis* 2nd Ed.,
5 Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate,
10 -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and
15 convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to
20 another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred
25 embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

30 When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence

similarity” comprises conservative or moderately conservative amino acid substitutions. A “conservative amino acid substitution” is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. See, e.g., Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A “moderately conservative” replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as “Gap” and “Bestfit” which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutin thereof. See, e.g., GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. See, e.g., Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein

5 incorporated by reference. Preferred parameters for blastp are:

	Expectation value:	10 (default)
	Filter:	seg (default)
	Cost to open a gap:	11 (default)
	Cost to extend a gap:	1 (default)
10	Max. alignments:	100 (default)
	Word size:	11 (default)
	No. of descriptions:	100 (default)
	Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at
15 least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms
20 other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (e.g., FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be
25 determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular
30 species, e.g., a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv,

dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

10 By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

15 A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but
20 using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it
25 an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one
30 or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For

instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than $1 \mu\text{M}$, preferably less than 100 nM and most preferably less than 10 nM .

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "colon specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the colon as compared to other tissues in the body. In a preferred embodiment, a "colon specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "colon specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more

- preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels,
- 5 such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

- 10 One aspect of the invention provides isolated nucleic acid molecules that are specific to the colon or to colon cells or tissue or that are derived from such nucleic acid molecules. These isolated colon specific nucleic acids (CSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid
- 15 molecule encodes a polypeptide that is specific to colon, a colon-specific polypeptide (CSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 148 through 250. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 147.

- 20 A CSNA may be derived from a human or from another animal. In a preferred embodiment, the CSNA is derived from a human or other mammal. In a more preferred embodiment, the CSNA is derived from a human or other primate. In an even more preferred embodiment, the CSNA is derived from a human.

- By "nucleic acid molecule" for purposes of the present invention, it is also meant
- 25 to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a CSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a CSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a CSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively
- 30 hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 148 through 250. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 147.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a CSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 148 through 250. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 147. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a CSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human CSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a CSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 148 through 250, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a CSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a CSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a CSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 147. In a preferred
5 embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a CSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 147, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a CSNA, more preferably at least 95%,
10 more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a CSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one
15 that exhibits sequence identity over its entire length to a CSNA or to a nucleic acid molecule encoding a CSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the CSNA or the nucleic acid molecule encoding a CSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at
20 least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 148 through 250 or demonstrates
25 significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 147. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the CSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated
30 species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived

from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In
5 another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a CSNA. Further, the substantially similar nucleic acid molecule may or may not be a CSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a CSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a
10 CSNA or a nucleic acid encoding a CSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population.
15 Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic
20 variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a CSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a CSNA. In a more preferred embodiment, the gene
25 is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 147. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic
30 acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a CSP. However, in a preferred embodiment, the part encodes a CSP. In one aspect, the invention comprises a part of a

CSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a CSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a CSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a CSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutin or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or

fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as ^{-32}P -dATP, ^{-32}P -dCTP, ^{-32}P -dGTP, ^{-32}P -dTTP, ^{-32}P -3'dATP, ^{-32}P -ATP, ^{-32}P -CTP, ^{-32}P -GTP, ^{-32}P -UTP, ^{-35}S -dATP, α - ^{35}S -GTP, α - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for

RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see Alers et al., Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al., J. NIH Res.* 5: 82 (1994); Van Belkum *et al., BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g., Tyagi et al., Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al., Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al., Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al., Science* 279: 1228-1229 (1998); Marras *et al., Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al., Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al., Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al., Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside

bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation,

10 phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having

15 normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019;

20 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

25 Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a

30 nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and

methylenedihydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average).

Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.,*

- 5 Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure

- 10 throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

- Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, 20 partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by 25 reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a CSNA, such as deletions, insertions, translocations, and duplications of the CSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g., Andreeff et al. (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify CSNA in, and isolate CSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g., Schwarchzacher et al., In Situ Hybridization*, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing

nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to CSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the
5 nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in
10 their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a CSP. In a more
15 preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 148 through 250. In another preferred embodiment, the probe or primer is derived from a CSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147.

20 In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer
25 in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide
30 probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter*

- alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

- The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

- Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as

hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes
5 (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

10 Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui
15 (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

20 Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of
25 this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for
30 example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

- In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the
- 5 nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*,
- 10 M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.
- 15 In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components,
- 20 *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series
- 25 plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most
- 30 common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

-47-

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway.

Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, e.g., in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A).

Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (e.g., vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (e.g., bovine papillomavirus), and retroviral vectors (e.g., murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast _-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the

5 transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the
10 Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the CSNA of interest. Often, expression is enhanced by
15 incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker
20 gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted
25 in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product
30 information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity

resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like

chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. See Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. See Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (see, e.g. Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (see, e.g. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (see, e.g. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (see, e.g. Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. See also Conn (ed.), Green Fluorescent Protein (Methods in

Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

5 Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

10 For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin
15 C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney
20 murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency;
25 varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art
30 may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must

be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present
5 invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,
10 and acylation, and it is an aspect of the present invention to provide CSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation,
15 acetylation, prenylation, methylation, acetylation, arginylation, ubiquitination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational
20 modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-
25 anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in
30 web sites such as the Delta Mass database <http://www.abrf.org/ABRF/ResearchCommittees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological

sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and
<http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a
revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27:
370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October
5 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu
et al. *Nucleic Acids Res* 27(1):237-239 (1999) and [http://www.cbs.dtu.dk/
databases/PhosphoBase/](http://www.cbs.dtu.dk/databases/PhosphoBase/) (accessed October 19, 2001); or [http://pir.georgetown.edu/
pirwww/search/textresid.html](http://pir.georgetown.edu/pirwww/search/textresid.html) (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational
10 modifications of proteins. Thus, in another embodiment, the invention provides
polypeptides from cancerous cells or tissues that have altered post-translational
modifications compared to the post-translational modifications of polypeptides from
normal cells or tissues. A number of altered post-translational modifications are known.
One common alteration is a change in phosphorylation state, wherein the polypeptide
15 from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated
compared to the polypeptide from a normal tissue, or wherein the polypeptide is
phosphorylated on different residues than the polypeptide from a normal cell. Another
common alteration is a change in glycosylation state, wherein the polypeptide from the
cancerous cell or tissue has more or less glycosylation than the polypeptide from a
20 normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a
different type of glycosylation than the polypeptide from a noncancerous cell or tissue.
Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-
carbohydrate interactions are important in cancer cell progression, dissemination and
invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer*
25 *Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is
prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either
farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a
protein to a cell membrane and is often required for polypeptide function. For instance,
30 the Ras superfamily of GTPase signaling proteins must be prenylated for function in a
cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et
al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to
5 the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is
10 cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell
15 compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method
20 known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel
25 electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the
30 art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically.

- 5 Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule
- 10 encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired
- 15 post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid
- 20 sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

- In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its
- 25 controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture
- 30 requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly
5 expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors
10 typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation,
15 liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector
20 comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to
25 express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture.
30 Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces*

pombe, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical

5 mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

10 readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from colon are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human colon cells.

15 Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel

20 (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

25 Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or

30 electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80

(1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) (<http://www.biorad.com/LifeScience/pdf/>

10 New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by

means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

5 Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a colon specific polypeptide (CSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ
10 ID NO: 148 through 250. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide,
15 wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a CSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 148 through 250. A polypeptide that comprises only a fragment of an entire CSP may or may not be a polypeptide that is also a CSP. For instance, a full-length polypeptide may be
20 colon-specific, while a fragment thereof may be found in other tissues as well as in colon. A polypeptide that is not a CSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-CSP antibodies. However, in a preferred embodiment, the part or fragment is a CSP. Methods of determining whether a polypeptide is a CSP are described *infra*.

25 Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by
30 native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. See, e.g., Lerner, *Nature* 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

10 Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

20 The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

25 One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, e.g., a CSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. See, e.g., Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably a CSP, may be produced by

30

chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a CSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be colon-specific. In a preferred embodiment, the mutein is colon-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 148 through 250. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino

acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is colon-specific, as described below. Multiple random mutations can be introduced into the

5 gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*,

10 Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the

15 polypeptide is homologous to a CSP. In an even more preferred embodiment, the polypeptide is homologous to a CSP selected from the group having an amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a CSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity

20 to an comprising an amino acid sequence of SEQ ID NO: 148 through 250. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In a yet more

25 preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%,

30 99.7%, 99.8% or 99.9% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a CSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a CSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the CSNA is selected from the group consisting of SEQ ID NO: 1 through 147. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a CSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the CSP is selected from the group consisting of SEQ ID NO: 148 through 250.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 148 through 250. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the CSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a CSP. Further, the homologous protein may or may not encode polypeptide that is a CSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a CSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but
5 also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in
10 eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a CSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino
15 acid sequence selected from the group consisting of SEQ ID NO: 148 through 250. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 147.

In another embodiment, the invention provides polypeptides which comprise
20 derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a CSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 148 through 250, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been
25 acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

30 Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of

glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.),

5 Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of

10 ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide,

15 including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic

20 processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on

25 the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546,

30 Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa

Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

- 10 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

- 25 The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

- 30 The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-CSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 10 In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a CSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 148 through 250. In a preferred embodiment, the analog is one that comprises one or
- 15 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a CSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
- 20 --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a CSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 25 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 30 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are

- described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

- Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and *t*BOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

- Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

- A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-

trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-
5 2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-
10 hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-
15 methylbenzoic acid, Fmoc-3-amino-2-naphtic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyl-dopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-
20 1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed
25 mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*,
30 *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a CSP. In a more preferred embodiment, the polypeptide
5 that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 148 through 250, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 147, or comprises all or part of a nucleic acid sequence that
10 selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 147.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of
15 the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present
20 invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the
25 present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically
30 synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so

included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbriizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95,

14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A
5 comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli*-LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

- 10 Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be
15 fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, α -amylase,
20 maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast ϕ -mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. See, e.g., Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may
25 also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (e.g., a Merrifield synthesis), or produced by chemical cross-linking.

- 30 Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the CSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize CSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to
5 assay for the polypeptides of the present invention, particularly CSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of CSPs, as for example by immunoprecipitation, and for use
10 as specific agonists or antagonists of CSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine
15 scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are
20 available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of
25 one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of
30 absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

-77-

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens
5 to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes; acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present
10 invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically
15 derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate,
20 such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when
25 the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with
30 sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so

-78-

attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

5

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a CSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 148 through 250, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a CSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or *visa versa*. In addition, alternative splice forms of a CSP may be indicative of cancer. Differential degradation of the C or N-terminus of a CSP may also be a marker or target for anticancer therapy. For example, a CSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-CSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine

the presence of the protein of the present invention in samples derived from human colon.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but

also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are
5 hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids
10 of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments
15 of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725
20 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation
25 and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant,
30 and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in

immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues

5 (Viking et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody

10 Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal

15 antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two

20 necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

25 Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies

30 of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein

- (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.,* Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997);
- 5 Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).
- Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas
- 10 (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

- Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length
- 15 antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

- For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al.*, *Biosci.*
- 20 *Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3): 157-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein
- 25 by reference in their entireties.

- Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997);
- 30 and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,
5 *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present
10 invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody
15 fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

20 Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

25 The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and
30 protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that
5 bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or
10 one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half
15 life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.,* United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21):
20 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region,
25 Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody
30 complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BlueGal; iodonitrotetrazolium

(INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are
5 luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic
10 compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such
15 enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

20 There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein
25 (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc.,
30 Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY

-87-

630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, 5 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting 10 applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , 15 ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

20 As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the 25 antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

30 The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention,

-88-

to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a CSP. In a preferred embodiment, the CSP comprises an amino acid sequence selected from SEQ ID NO: 148 through 250, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a CSNA of the invention, preferably a CSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 147, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human CSG.

The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual*, 2d ed., Cold Spring Harbor Press (1999); Jackson et al., *Mouse Genetics and Transgenics: A Practical Approach*, Oxford University Press (2000); and Pinkert, *Transgenic Animal Technology: A Laboratory Handbook*, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g., Paterson et al., Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al., *Biotechnology* 11: 1263-1270 (1993); Wright et al., *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989) retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g., Ulmer et al., Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem

cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g., Lavitrano et al., Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,*
5 *Campell et al., Nature* 380: 64-66 (1996); *Wilmot et al., Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a*
nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i. e.,* mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such
10 as in concatamers, *e. g.,* head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by
following, *e.g.,* the teaching of Lasko *et al. et al., Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation
will depend upon the particular cell type of interest, and will be apparent to those of skill
15 in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression
20 of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the
25 transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in
30 order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to

both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

5 Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

10 Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous
15 gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable
25 marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications
30 to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are

directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

-93-

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 147 and SEQ ID NO: 148 through 250 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data

set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences
5 comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or
10 data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and
15 similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a
20 nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence
25 comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the
30 steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one

overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Colon Cancer

- 5 The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a CSNA or a CSP in a human patient that has or may have colon cancer, or who is at risk of developing colon cancer, with the expression of a CSNA or a CSP in a normal human control. For purposes of the present invention,
- 10 “expression of a CSNA” or “CSNA expression” means the quantity of CSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term “expression of a CSP” or “CSP expression” means the amount of CSP that can be measured by any method known in the art or the level of translation of a CSG
- 15 CSNA that can be measured by any method known in the art.

- The present invention provides methods for diagnosing colon cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of CSNA or CSP in cells, tissues, organs or bodily fluids compared with levels of CSNA or CSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human
- 20 control, wherein an increase, or decrease in certain cases, in levels of a CSNA or CSP in the patient versus the normal human control is associated with the presence of colon cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in the structure of the mRNA of a CSG compared to the mRNA from a normal
- 25 control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in a CSP compared to a CSP from a normal control. These changes include, e.g., alterations in glycosylation and/or phosphorylation of the CSP or
- 30 subcellular CSP localization.

 In a preferred embodiment, the expression of a CSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from

SEQ ID NO: 148 through 250, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the CSNA expression that is measured is the level of expression of a CSNA mRNA selected from SEQ ID NO: 1 through 147, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of
5 any of these nucleic acids. CSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.,* Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. CSNA
10 transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a CSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.,* aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, CSNA expression may be compared to a known control, such as
15 normal colon nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a CSP is measured by determining the level of a CSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 148 through 250, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs
20 and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of CSNA or CSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of colon cancer. The expression level of a CSP may be determined by any method known in the art, such as those
25 described *supra*. In a preferred embodiment, the CSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.,* Harlow (1999),
30 *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the CSP structure may be determined by any method known in the art, including, *e.g.,* using antibodies that specifically recognize phosphoserine, phosphothreonine or

-97-

phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a CSP is prepared if one is not already available. In a preferred
5 embodiment, the antibody is a monoclonal antibody. The anti-CSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the CSP will bind to the anti-CSP antibody. The sample is removed, the solid support is washed to remove unbound
10 material, and an anti-CSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the CSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the
15 amount of a CSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure CSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-CSP antibody is attached to a solid
20 support and an allocated amount of a labeled CSP and a sample of interest are incubated with the solid support. The amount of labeled CSP detected which is attached to the solid support can be correlated to the quantity of a CSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential
25 separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly
30 square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a CSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example,

5 reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

10 Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more CSNAs of interest. In this approach, all or a portion of one or more CSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the
15 RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid
20 molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any
25 other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of CSNA or CSP includes, without limitation, colon tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, colon cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment,
30 especially when metastasis of a primary colon cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without

limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in
5 CSNAs or CSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a CSNA or CSP. In many cases, the use of another cancer marker
10 will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other CSNA or CSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular CSNA or CSP is measured. In a more
15 preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression
20 levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having colon cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP and then ascertaining whether the patient has colon cancer from the expression level of the CSNA or CSP. In general, if high expression
25 relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is
30 indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the

same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether colon cancer has metastasized in a patient. One may identify whether the colon cancer has metastasized by measuring the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a variety of tissues. The presence of a CSNA or CSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a CSNA or CSP is associated with colon cancer. Similarly, the presence of a CSNA or CSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a CSNA or CSP is associated with colon cancer. Further, the presence of a structurally altered CSNA or CSP that is associated with colon cancer is also indicative of metastasis.

In general, if high expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The CSNA or CSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with colon cancers or other colon related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of colon disorders.

30 *Staging*

The invention also provides a method of staging colon cancer in a human patient. The method comprises identifying a human patient having colon cancer and analyzing

-101-

cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more CSNAs or CSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more CSNAs or CSPs is determined for each stage to obtain a standard expression level for each CSNA and CSP. Then, the CSNA or CSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The CSNA or CSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the CSNAs and CSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a CSNA or CSP to determine the stage of a colon cancer.

Monitoring

Further provided is a method of monitoring colon cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the colon cancer. The method comprises identifying a human patient that one wants to monitor for colon cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more CSNAs or CSPs, and comparing the CSNA or CSP levels over time to those CSNA or CSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a CSNA or CSP that are associated with colon cancer.

If increased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a

cancerous lesion, then detecting an decrease in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of CSNAs or CSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

- 5 Monitoring a patient for onset of colon cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a CSNA and/or CSP. The present invention
10 provides a method in which a test sample is obtained from a human patient and one or more CSNAs and/or CSPs are detected. The presence of higher (or lower) CSNA or CSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly colon cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more CSNAs and/or
15 CSPs of the invention can also be monitored by analyzing levels of expression of the CSNAs and/or CSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

20 *Detection of Genetic Lesions or Mutations*

The methods of the present invention can also be used to detect genetic lesions or mutations in a CSG, thereby determining if a human with the genetic lesion is susceptible to developing colon cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing colon cancer. Genetic lesions can be detected,
25 for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the CSGs of this invention, a chromosomal rearrangement of CSG, an aberrant modification of CSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a CSG. Methods to detect such lesions in the CSG of this invention are known to those having ordinary skill in the art following the teachings
30 of the specification.

Methods of Detecting Noncancerous Colon Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having or known to have a noncancerous colon disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP, comparing the expression level or structural alteration of the CSNA or CSP to a normal colon control, and then ascertaining whether the patient has a noncancerous colon disease. In general, if high expression relative to a control of a CSNA or CSP is indicative of a particular noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of a noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a CSNA and/or CSP is associated with a particular noncancerous colon disease by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining which CSNAs and/or CSPs are expressed in the tissue at either a higher or a lower level than in normal colon tissue. In another embodiment, one may determine whether a CSNA or CSP exhibits structural alterations in a particular noncancerous colon disease state by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining the structural alterations in one or more CSNAs and/or CSPs relative to normal colon tissue.

Methods for Identifying Colon Tissue

In another aspect, the invention provides methods for identifying colon tissue.

These methods are particularly useful in, *e.g.*, forensic science, colon cell differentiation
5 and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is colon tissue or has colon tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising colon tissue or having colon tissue-like characteristics, determining whether the sample expresses one or more CSNAs
10 and/or CSPs, and, if the sample expresses one or more CSNAs and/or CSPs, concluding that the sample comprises colon tissue. In a preferred embodiment, the CSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 148 through 250, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 147, or a
15 hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a CSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a CSP is expressed. Determining whether a sample
20 expresses a CSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the CSP has an amino acid sequence selected from SEQ ID NO: 148 through 250, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two CSNAs and/or CSPs is determined. In a more preferred embodiment, the
25 expression of at least three, more preferably four and even more preferably five CSNAs and/or CSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is colon tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are
30 recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into colon tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new colon tissue by tissue engineering. These agents include,

-105-

e.g., growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

5 Methods for Producing and Modifying Colon Tissue

In another aspect, the invention provides methods for producing engineered colon tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a CSNA or a CSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of colon tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal colon tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered colon tissue or cells comprises one of these cell types. In another embodiment, the engineered colon tissue or cells comprises more than one colon cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the colon cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more CSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode CSPs having amino acid sequences selected from SEQ ID NO: 148 through 250, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 147, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a CSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial colon tissue may be used to treat patients who have lost some or all of their colon function.

Pharmaceutical Compositions

30

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives,

antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a CSNA or part thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 147, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a CSP or fragment thereof. In a more preferred embodiment, the CSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 148 through 250, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-CSP antibody, preferably an antibody that specifically binds to a CSP having an amino acid that is selected from the group consisting of SEQ ID NO: 148 through 250, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or

suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

5 Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing,
10 together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

15 For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum
20 albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and
25 administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate,
30 isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of
5 the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature
10 of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered
15 topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include
20 chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from
25 about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be
30 administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for
5 treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the
10 present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an
15 appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for
20 example CSP polypeptide, fusion protein, or fragments thereof, antibodies specific for CSP, agonists, antagonists or inhibitors of CSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present
25 invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the
30 population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture or animal model systems. The dose ratio of toxic to therapeutic

-111-

effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of
5 circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

10 The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting
15 pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody
20 agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will
25 employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to
30 the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of colon function. As used herein, “treating” includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term “treating” encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

10 *Gene Therapy and Vaccines*

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of “naked” nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a CSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a CSP are administered, for example, to complement a deficiency in the native CSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid

-113-

molecule encodes a CSP having the amino acid sequence of SEQ ID NO: 148 through 250, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a CSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in CSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a CSP having the amino acid sequence of SEQ ID NO: 148 through 250, or a fragment, fusion protein, allelic variant or homolog thereof.

10 *Antisense Administration*

Antisense nucleic acid compositions, or vectors that drive expression of a CSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a CSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

15 Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a CSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to CSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

25 Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the CSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a CSP, preferably a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule
5 having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a CSP, a
10 fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant CSP defect.

Protein compositions are administered, for example, to complement a deficiency in native CSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to CSP. The immune response can
15 be used to modulate activity of CSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate CSP.

In a preferred embodiment, the polypeptide is a CSP comprising an amino acid
20 sequence of SEQ ID NO: 148 through 250, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example,
30 to antagonize activity of CSP, or to target therapeutic agents to sites of CSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a

-115-

CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic
5 variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a CSP or have a modulatory effect on the expression or activity of a CSP. Modulators which decrease the expression or activity of CSP (antagonists) are believed to be useful in treating colon cancer. Such screening assays are known to those of skill in
10 the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a CSP can also be designed, synthesized and tested for use in the imaging and treatment of colon cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the CSPs identified herein. Molecules
15 identified in the library as being capable of binding to a CSP are key candidates for further evaluation for use in the treatment of colon cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a CSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of CSP is
20 administered. Antagonists of CSP can be produced using methods generally known in the art. In particular, purified CSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a CSP.

In other embodiments a pharmaceutical composition comprising an agonist of a
25 CSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP comprising an amino acid sequence of SEQ ID NO: 148 through 250, or a fusion protein, allelic variant, homolog, analog or
30 derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP encoded by a

nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 147, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Colon Tissue

- The invention also provides a method in which a polypeptide of the invention, or
- 5 an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the colon or to specific cells in the colon. In a preferred embodiment, an anti-CSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if colon tissue needs to be selectively destroyed. This would be useful for targeting and killing colon cancer cells.
- 10 In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting colon cell function.

- In another embodiment, an anti-CSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring colon function, identifying colon cancer tumors,
- 15 and identifying noncancerous colon diseases.

EXAMPLES

Example 1: Gene Expression analysis

- CSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using
- 20 the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression
- 25 in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a
- 30 simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer"

disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be
5 considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the CSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4);
10 differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into “gene bins,” where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with
15 rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie “The significance of digital gene expression profiles,” Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are
20 incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were
25 analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

- (a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a
30 gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.

- (b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.
- (c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

The CLASP™ scores for SEQ ID NO: 1-147 are listed below:

	SEQ ID NO: 1	DEX0253_1	CLASP2
	SEQ ID NO: 2	DEX0253_2	CLASP2
15	SEQ ID NO: 3	DEX0253_3	CLASP2
	SEQ ID NO: 4	DEX0253_4	CLASP2 CLASP1
	SEQ ID NO: 5	DEX0253_5	CLASP2 CLASP1
	SEQ ID NO: 6	DEX0253_6	CLASP5 CLASP1
	SEQ ID NO: 7	DEX0253_7	CLASP5 CLASP1
20	SEQ ID NO: 8	DEX0253_8	CLASP2
	SEQ ID NO: 9	DEX0253_9	CLASP2
	SEQ ID NO: 10	DEX0253_10	CLASP2
	SEQ ID NO: 11	DEX0253_11	CLASP2 CLASP1
	SEQ ID NO: 12	DEX0253_12	CLASP2
25	SEQ ID NO: 13	DEX0253_13	CLASP2
	SEQ ID NO: 14	DEX0253_14	CLASP5 CLASP1
	SEQ ID NO: 15	DEX0253_15	CLASP2
	SEQ ID NO: 16	DEX0253_16	CLASP2
	SEQ ID NO: 17	DEX0253_17	CLASP2
30	SEQ ID NO: 18	DEX0253_18	CLASP2
	SEQ ID NO: 19	DEX0253_19	CLASP2
	SEQ ID NO: 20	DEX0253_20	CLASP2
	SEQ ID NO: 21	DEX0253_21	CLASP2
	SEQ ID NO: 22	DEX0253_22	CLASP2
35	SEQ ID NO: 23	DEX0253_23	CLASP2
	SEQ ID NO: 24	DEX0253_24	CLASP2 CLASP1
	SEQ ID NO: 25	DEX0253_25	CLASP2 CLASP1
	SEQ ID NO: 26	DEX0253_26	CLASP2 CLASP1
	SEQ ID NO: 27	DEX0253_27	CLASP2 CLASP1
40	SEQ ID NO: 28	DEX0253_28	CLASP2
	SEQ ID NO: 29	DEX0253_29	CLASP2
	SEQ ID NO: 30	DEX0253_30	CLASP2

-119-

	SEQ ID NO: 31	DEX0253_31 CLASP2
	SEQ ID NO: 32	DEX0253_32 CLASP2
	SEQ ID NO: 33	DEX0253_33 CLASP2
	SEQ ID NO: 34	DEX0253_34 CLASP2
5	SEQ ID NO: 35	DEX0253_35 CLASP2
	SEQ ID NO: 36	DEX0253_36 CLASP2
	SEQ ID NO: 37	DEX0253_37 CLASP2
	SEQ ID NO: 38	DEX0253_38 CLASP2
	SEQ ID NO: 39	DEX0253_39 CLASP2
10	SEQ ID NO: 40	DEX0253_40 CLASP2
	SEQ ID NO: 41	DEX0253_41 CLASP2
	SEQ ID NO: 42	DEX0253_42 CLASP2
	SEQ ID NO: 43	DEX0253_43 CLASP2
	SEQ ID NO: 44	DEX0253_44 CLASP2
15	SEQ ID NO: 45	DEX0253_45 CLASP2
	SEQ ID NO: 46	DEX0253_46 CLASP2
	SEQ ID NO: 47	DEX0253_47 CLASP2
	SEQ ID NO: 48	DEX0253_48 CLASP2 CLASP1
	SEQ ID NO: 50	DEX0253_50 CLASP2 CLASP1
20	SEQ ID NO: 51	DEX0253_51 CLASP2 CLASP1
	SEQ ID NO: 52	DEX0253_52 CLASP2 CLASP1
	SEQ ID NO: 53	DEX0253_53 CLASP2
	SEQ ID NO: 54	DEX0253_54 CLASP2 CLASP1
	SEQ ID NO: 55	DEX0253_55 CLASP2 CLASP1
25	SEQ ID NO: 56	DEX0253_56 CLASP2
	SEQ ID NO: 57	DEX0253_57 CLASP2
	SEQ ID NO: 58	DEX0253_58 CLASP2
	SEQ ID NO: 59	DEX0253_59 CLASP2
	SEQ ID NO: 60	DEX0253_60 CLASP2
30	SEQ ID NO: 61	DEX0253_61 CLASP2
	SEQ ID NO: 62	DEX0253_62 CLASP5 CLASP1
	SEQ ID NO: 63	DEX0253_63 CLASP5 CLASP1
	SEQ ID NO: 64	DEX0253_64 CLASP5 CLASP1
	SEQ ID NO: 65	DEX0253_65 CLASP5 CLASP1
35	SEQ ID NO: 66	DEX0253_66 CLASP2
	SEQ ID NO: 68	DEX0253_68 CLASP2
	SEQ ID NO: 69	DEX0253_69 CLASP2 CLASP1
	SEQ ID NO: 70	DEX0253_70 CLASP5 CLASP1
	SEQ ID NO: 71	DEX0253_71 CLASP5 CLASP1
40	SEQ ID NO: 72	DEX0253_72 CLASP2
	SEQ ID NO: 73	DEX0253_73 CLASP2
	SEQ ID NO: 74	DEX0253_74 CLASP2
	SEQ ID NO: 75	DEX0253_75 CLASP2
	SEQ ID NO: 76	DEX0253_76 CLASP2 CLASP1 CLASP4
45	SEQ ID NO: 77	DEX0253_77 CLASP2
	SEQ ID NO: 78	DEX0253_78 CLASP5 CLASP1
	SEQ ID NO: 79	DEX0253_79 CLASP5 CLASP1
	SEQ ID NO: 80	DEX0253_80 CLASP5 CLASP1

	SEQ ID NO: 81	DEX0253_81	CLASP5 CLASP1
	SEQ ID NO: 82	DEX0253_82	CLASP2
	SEQ ID NO: 83	DEX0253_83	CLASP2
	SEQ ID NO: 84	DEX0253_84	CLASP1
5	SEQ ID NO: 85	DEX0253_85	CLASP1
	SEQ ID NO: 86	DEX0253_86	CLASP2
	SEQ ID NO: 87	DEX0253_87	CLASP2
	SEQ ID NO: 88	DEX0253_88	CLASP2
	SEQ ID NO: 89	DEX0253_89	CLASP2
10	SEQ ID NO: 90	DEX0253_90	CLASP2
	SEQ ID NO: 91	DEX0253_91	CLASP2
	SEQ ID NO: 92	DEX0253_92	CLASP2
	SEQ ID NO: 93	DEX0253_93	CLASP2
	SEQ ID NO: 94	DEX0253_94	CLASP2
15	SEQ ID NO: 95	DEX0253_95	CLASP2 CLASP1
	SEQ ID NO: 96	DEX0253_96	CLASP2 CLASP1
	SEQ ID NO: 97	DEX0253_97	CLASP2
	SEQ ID NO: 98	DEX0253_98	CLASP2
	SEQ ID NO: 99	DEX0253_99	CLASP2
20	SEQ ID NO: 100	DEX0253_100	CLASP2
	SEQ ID NO: 101	DEX0253_101	CLASP2
	SEQ ID NO: 102	DEX0253_102	CLASP2
	SEQ ID NO: 103	DEX0253_103	CLASP2
	SEQ ID NO: 104	DEX0253_104	CLASP5 CLASP1
25	SEQ ID NO: 105	DEX0253_105	CLASP2
	SEQ ID NO: 106	DEX0253_106	CLASP2
	SEQ ID NO: 107	DEX0253_107	CLASP2
	SEQ ID NO: 108	DEX0253_108	CLASP2
	SEQ ID NO: 109	DEX0253_109	CLASP2 CLASP1
30	SEQ ID NO: 110	DEX0253_110	CLASP2
	SEQ ID NO: 111	DEX0253_111	CLASP2
	SEQ ID NO: 112	DEX0253_112	CLASP2
	SEQ ID NO: 113	DEX0253_113	CLASP2
	SEQ ID NO: 114	DEX0253_114	CLASP2
35	SEQ ID NO: 115	DEX0253_115	CLASP5 CLASP1
	SEQ ID NO: 116	DEX0253_116	CLASP2
	SEQ ID NO: 117	DEX0253_117	CLASP2
	SEQ ID NO: 118	DEX0253_118	CLASP2
	SEQ ID NO: 119	DEX0253_119	CLASP2
40	SEQ ID NO: 120	DEX0253_120	CLASP2
	SEQ ID NO: 121	DEX0253_121	CLASP2
	SEQ ID NO: 123	DEX0253_123	CLASP2
	SEQ ID NO: 124	DEX0253_124	CLASP2
	SEQ ID NO: 125	DEX0253_125	CLASP5 CLASP1
45	SEQ ID NO: 126	DEX0253_126	CLASP5 CLASP1
	SEQ ID NO: 127	DEX0253_127	CLASP2
	SEQ ID NO: 128	DEX0253_128	CLASP2
	SEQ ID NO: 129	DEX0253_129	CLASP2

-121-

	SEQ ID NO: 130	DEX0253_130	CLASP2
	SEQ ID NO: 131	DEX0253_131	CLASP2
	SEQ ID NO: 132	DEX0253_132	CLASP2
	SEQ ID NO: 133	DEX0253_133	CLASP5 CLASP1
5	SEQ ID NO: 134	DEX0253_134	CLASP2
	SEQ ID NO: 135	DEX0253_135	CLASP2
	SEQ ID NO: 136	DEX0253_136	CLASP2
	SEQ ID NO: 139	DEX0253_139	CLASP2
	SEQ ID NO: 140	DEX0253_140	CLASP5
10	SEQ ID NO: 141	DEX0253_141	CLASP2
	SEQ ID NO: 142	DEX0253_142	CLASP2
	SEQ ID NO: 143	DEX0253_143	CLASP2
	SEQ ID NO: 144	DEX0253_144	CLASP2
	SEQ ID NO: 145	DEX0253_145	CLASP2
15	SEQ ID NO: 146	DEX0253_146	CLASP2
	SEQ ID NO: 147	DEX0253_147	CLASP2

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation
 20 detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The
 method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5'
 reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity
 of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected
 by the laser detector of the Model 7700 Sequence Detection System (PE Applied
 25 Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to
 standardize the amount of sample RNA added to the reaction and normalize for Reverse
 Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate
 dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this
 endogenous control. To calculate relative quantitation between all the samples studied,
 30 the target RNA levels for one sample were used as the basis for comparative results
 (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard
 curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence
 Detection System).

The tissue distribution and the level of the target gene are evaluated for every
 35 sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer
 tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently,
 first strand cDNA is prepared with reverse transcriptase and the polymerase chain
 reaction is done using primers and Taqman probes specific to each target gene. The

results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the CSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the CSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the CSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 147 being a diagnostic marker for cancer.

DEX0095_71

25 **cln150-sqcln033**

Sequence

Dex0095_71 (cln150-sqcln033)

Sequence ID #

DEX0253_104(SEQ ID No.:104)

Semi quantitative PCR was done using the following primers:

30

Primer	DexSeqID	From	To	Primer Length
Sqcln033F	DEX0253_104	36	56	21
Sqcln033R	DEX0253_104	256	236	21

-123-

The relative levels of expression of sqcln0033 in 12 normal samples from 12 different tissues were determined. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Normal
Breast	0
Colon	1000
Endometrium	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Prostate	0
Small Intestine	1000
Stomach	0
Testis	10
Uterus	0

10 Relative levels of expression in the table above show that none of the normal tissues examined here except colon and small intestine exhibit appreciable levels of sqcln033 gene expression.

The relative levels of expression of sqcln033 in 12 cancer samples from 12 different tissues were determined. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Cancer
Bladder	1
Breast	0
Colon	1000
Kidney	1
Liver	0
Lung	0
Ovary	0
pancreas	0

-124-

prostate	1
Stomach	0
Testes	1
Uterus	0

Relative levels of expression in Table 2 show that sqcln033 is expressed in very low levels in bladder, kidney, prostate and testicular cancer. High levels of expression were observed for colon cancer.

- 5 The relative levels of expression of sqcln033 in 6 matching pair colon samples were determined. Each matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

- 10 Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Sample ID	Tissue	Cancer	NAT
9609B019	Colon	10	100
9709C074RA	Colon	1	10
9705F002D	Colon	10	100
9608B012	Colon	0	100
4004709A1	Colon	10	10
9707C004GB	Colon	100	10

- 15 Relative levels of expression in the table above show that sqcln033 is expressed in low levels in four and in moderate levels in one of the six colon cancer samples. Moderate levels of expression was observed in matching normal adjacent tissue (NAT).

- 20 From the semi-quantitative PCR experiments exhibited above especially from the results in the tables above high degree of specificity of sqCln033 for colon tissue is observed.

Experiments are underway to design and test primers and probe for quantitative PCR experiments.

Example 3: Protein Expression

- 25 The CSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the CSNA is subcloned in pET-21d for expression in *E. coli*. In

addition to the CSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of CSNA, and six histidines, flanking the COOH-terminus of the coding sequence of CSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

- 5 An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

 Large-scale purification of CSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated
10 with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. CSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

- 15 Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI
20 cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring
25 signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.,* WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

- 30 In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such

cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The

5 splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*,

10 *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes

15 use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability

20 to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by

25 reference).

	DEX0253_151	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	68-108	1.02	41
30	DEX0253_154	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	13-31	1.00	19
	DEX0253_155	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	8-22	1.25	15
35	67-98	1.06	32

-127-

	DEX0253_156	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	14-33	1.13 20
	DEX0253_161	Antigenicity Index(Jameson-Wolf)
5	positions	AI avg length
	32-46	0.98 15
	DEX0253_170	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	6-48	1.02 43
10	DEX0253_172	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	19-28	1.00 10
	DEX0253_174	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
15	13-29	1.01 17
	DEX0253_178	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	2-17	0.99 16
	DEX0253_179	Antigenicity Index(Jameson-Wolf)
20	positions	AI avg length
	7-20	0.96 14
	DEX0253_184	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	18-33	1.00 16
25	DEX0253_187	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	42-54	1.09 13
	DEX0253_188	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
30	14-41	1.03 28
	81-97	0.96 17
	DEX0253_192	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	25-75	0.97 51
35	DEX0253_197	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	31-58	1.06 28
	DEX0253_198	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
40	73-82	1.22 10
	154-165	1.21 12
	213-224	1.06 12
	38-58	1.05 21
	314-323	1.03 10
45	250-267	0.99 18
	275-295	0.98 21
	137-149	0.96 13
	DEX0253_203	Antigenicity Index(Jameson-Wolf)

-128-

	positions	AI	avg length
	3-28	0.99	26
	DEX0253_204	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
5	29-39	1.10	11
	52-65	0.97	14
	DEX0253_208	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	3-14	1.04	12
10	DEX0253_211	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	33-54	1.22	22
	59-76	1.08	18
	DEX0253_212	Antigenicity Index(Jameson-Wolf)	
15	positions	AI	avg length
	22-31	1.00	10
	DEX0253_214	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	7-22	1.24	16
20	DEX0253_217	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	34-46	1.06	13
	DEX0253_218	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
25	23-32	1.16	10
	DEX0253_219	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	19-30	1.24	12
	DEX0253_221	Antigenicity Index(Jameson-Wolf)	
30	positions	AI	avg length
	58-69	1.08	12
	42-51	1.04	10
	DEX0253_222	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
35	58-69	1.08	12
	42-51	1.04	10
	DEX0253_226	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	8-18	1.01	11
40	DEX0253_229	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	8-36	1.02	29
	DEX0253_237	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
45	14-31	0.91	18
	DEX0253_245	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	3-27	1.12	25

-129-

DEX0253_246	Antigenicity Index(Jameson-Wolf)
positions	AI avg length
23-32	0.93 10

5 Examples of post-translational modifications (PTMs) of the BSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the LSPs

10 of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html most recently accessed October 23, 2001). For full definitions of the PTMs see <http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.

	DEX0253_148	Tyr_Phospho_Site 3-9;
	DEX0253_150	Ck2_Phospho_Site 19-22;23-26;
15	DEX0253_151	Ck2_Phospho_Site 83-86;128-131; Myristyl 28-33;49-54;
		Pkc_Phospho_Site 77-79;83-85;105-107;139-141;
	DEX0253_152	Leucine_Zipper 6-27; Pkc_Phospho_Site 16-18;
	DEX0253_153	Pkc_Phospho_Site 16-18;24-26;
	DEX0253_154	Ck2_Phospho_Site 2-5;23-26;
20	DEX0253_155	Amidation 12-15; Myristyl 12-17;31-36;34-39; Pkc_Phospho_Site 75-77;
	DEX0253_156	Ck2_Phospho_Site 8-11; Pkc_Phospho_Site 16-18;
	DEX0253_158	Pkc_Phospho_Site 2-4;
	DEX0253_159	Myristyl 26-31; Pkc_Phospho_Site 9-11;14-16;
25	DEX0253_160	Myristyl 13-18; Pkc_Phospho_Site 31-33;
	DEX0253_161	Pkc_Phospho_Site 8-10;17-19;
	DEX0253_162	Ck2_Phospho_Site 28-31; Myristyl 48-53; Pkc_Phospho_Site 28-30;
	DEX0253_164	Asn_Glycosylation 21-24; Myristyl 17-22;
30	DEX0253_165	Pkc_Phospho_Site 44-46;
	DEX0253_166	Asn_Glycosylation 7-10; Camp_Phospho_Site 3-6;
		Ck2_Phospho_Site 19-22; Myristyl 32-37; Pkc_Phospho_Site 6-8;45-47; Tyr_Phospho_Site 28-34;
	DEX0253_167	Asn_Glycosylation 14-17; Myristyl 23-28; Pkc_Phospho_Site 9-11;
35	DEX0253_168	Asn_Glycosylation 38-41; Ck2_Phospho_Site 40-43; Myristyl 49-54; Pkc_Phospho_Site 16-18;35-37;
	DEX0253_170	Camp_Phospho_Site 43-46; Myristyl 36-41;
	DEX0253_171	Asn_Glycosylation 37-40; Ck2_Phospho_Site 79-82;86-89;
40	DEX0253_172	Myristyl 22-27;34-39;47-52; Pkc_Phospho_Site 7-9;15-17;
		Asn_Glycosylation 7-10;11-14;21-24; Pkc_Phospho_Site 52-54;55-57;
	DEX0253_173	Ck2_Phospho_Site 27-30;
	DEX0253_174	Myristyl 56-61;

-130-

	DEX0253_175	Myristyl 24-29;
	DEX0253_176	Pkc_Phospho_Site 8-10;
	DEX0253_177	Pkc_Phospho_Site 9-11;
5	DEX0253_178	Ck2_Phospho_Site 52-55; Myristyl 24-29; Pkc_Phospho_Site 45-47;
	DEX0253_179	Asn_Glycosylation 63-66; Ck2_Phospho_Site 25-28;
		Pkc_Phospho_Site 27-29;
	DEX0253_180	Ck2_Phospho_Site 26-29;
	DEX0253_181	Amidation 51-54; Pkc_Phospho_Site 6-8;51-53;
10	DEX0253_183	Ck2_Phospho_Site 17-20; Pkc_Phospho_Site 36-38;39-41;
	DEX0253_184	Pkc_Phospho_Site 31-33;
	DEX0253_186	Pkc_Phospho_Site 18-20;
	DEX0253_187	Ck2_Phospho_Site 42-45;77-80;
	DEX0253_188	Ck2_Phospho_Site 21-24;86-89; Myristyl 25-30;
15		Pkc_Phospho_Site 73-75;
	DEX0253_189	Amidation 25-28; Camp_Phospho_Site 10-13; Ck2_Phospho_Site 37-40;
	DEX0253_191	Ck2_Phospho_Site 11-14; Myristyl 16-21;
	DEX0253_192	Ck2_Phospho_Site 48-51;61-64; Myristyl 57-62;
20		Pkc_Phospho_Site 22-24;58-60;
	DEX0253_193	Ck2_Phospho_Site 43-46; Pkc_Phospho_Site 18-20;34-36;
	DEX0253_196	Asn_Glycosylation 3-6; Ck2_Phospho_Site 8-11;
	DEX0253_197	Myristyl 4-9;
25	DEX0253_198	Asn_Glycosylation 46-49;54-57;99-102;123-126;135-138;182-185;195-198;223-226;229-232;272-275;305-308;
		Camp_Phospho_Site 259-262; Ck2_Phospho_Site 75-78;82-85;114-117;197-200;208-211;247-250;308-311;382-385;395-398;
		Myristyl 173-178;316-321;320-325;334-339;339-344;
		Pkc_Phospho_Site 63-65;82-84;101-103;114-116;157-159;187-189;197-199;212-214;225-227;293-295;395-397;
30		Receptor_Cytokines_1 126-138; Receptor_Cytokines_2 304-310;
	DEX0253_199	Camp_Phospho_Site 6-9;
	DEX0253_202	Amidation 38-41; Asn_Glycosylation 2-5;45-48;
	DEX0253_203	Camp_Phospho_Site 24-27; Pkc_Phospho_Site 20-22;
35	DEX0253_204	Asn_Glycosylation 65-68; Pkc_Phospho_Site 54-56;
	DEX0253_205	Ck2_Phospho_Site 6-9; Myristyl 46-51;47-52;59-64;63-68;78-83;
		Pkc_Phospho_Site 6-8;19-21;
	DEX0253_207	Ck2_Phospho_Site 57-60;61-64; Myristyl 69-74;
		Pkc_Phospho_Site 34-36;
40	DEX0253_208	Camp_Phospho_Site 5-8; Ck2_Phospho_Site 37-40; Myristyl 26-31;
	DEX0253_209	Pkc_Phospho_Site 52-54;
	DEX0253_210	Camp_Phospho_Site 34-37; Ck2_Phospho_Site 9-12;
		Pkc_Phospho_Site 9-11;33-35;
45	DEX0253_211	Amidation 67-70; Camp_Phospho_Site 69-72; Ck2_Phospho_Site 5-8;10-13;23-26; Myristyl 35-40;54-59; Pkc_Phospho_Site 67-69;
	DEX0253_212	Prokar_Lipoprotein 6-16;

-131-

	DEX0253_213	Asn_Glycosylation 44-47; Camp_Phospho_Site 37-40; Ck2_Phospho_Site 2-5; Pkc_Phospho_Site 8-10;
	DEX0253_215	Asn_Glycosylation 3-6; Ck2_Phospho_Site 28-31;
	DEX0253_216	Ck2_Phospho_Site 9-12; Pkc_Phospho_Site 32-34;
5	DEX0253_217	Camp_Phospho_Site 28-31; Myristyl 12-17; Rieske_2 35-40;
	DEX0253_218	Ck2_Phospho_Site 21-24; Myristyl 29-34;
	DEX0253_221	Camp_Phospho_Site 51-54; Ck2_Phospho_Site 29-32;98-101; Myristyl 21-26;70-75;75-80; Pkc_Phospho_Site 50-52;57-59;104- 106; Prokar_Lipoprotein 66-76;68-78;
10	DEX0253_222	Camp_Phospho_Site 51-54; Ck2_Phospho_Site 29-32;98-101; Myristyl 21-26;70-75;75-80; Pkc_Phospho_Site 50-52;57-59;104- 106; Prokar_Lipoprotein 66-76;68-78;
	DEX0253_223	Ck2_Phospho_Site 75-78;
	DEX0253_224	Ck2_Phospho_Site 96-99; Myristyl 72-77;76-81;105-110;
15		Pkc_Phospho_Site 63-65;87-89;
	DEX0253_225	Asn_Glycosylation 3-6;8-11;9-12; Pkc_Phospho_Site 12-14;
	DEX0253_226	Camp_Phospho_Site 14-17;
	DEX0253_228	Ck2_Phospho_Site 14-17;
	DEX0253_229	Asn_Glycosylation 41-44; Ck2_Phospho_Site 20-23;39-42;
20		Myristyl 25-30;30-35;37-42;45-50; Pkc_Phospho_Site 9-11;20-22;
	DEX0253_230	Ck2_Phospho_Site 15-18;
	DEX0253_231	Camp_Phospho_Site 25-28;
	DEX0253_232	Ck2_Phospho_Site 7-10; Rgd 8-10;
	DEX0253_235	Ck2_Phospho_Site 21-24; Myristyl 4-9; Pkc_Phospho_Site 21-23;
25	DEX0253_236	Ck2_Phospho_Site 12-15;26-29; Myristyl 10-15;24-29;30-35;
	DEX0253_237	Ck2_Phospho_Site 41-44;
	DEX0253_238	Ck2_Phospho_Site 2-5; Pkc_Phospho_Site 46-48;
	DEX0253_239	Pkc_Phospho_Site 9-11;17-19;
	DEX0253_240	Camp_Phospho_Site 2-5; Ck2_Phospho_Site 5-8;
30	DEX0253_241	Pkc_Phospho_Site 20-22;
	DEX0253_242	Ck2_Phospho_Site 30-33; Myristyl 4-9;27-32;36-41;
	DEX0253_243	Asn_Glycosylation 40-43; Pkc_Phospho_Site 12-14;48-50;
	DEX0253_245	Asn_Glycosylation 7-10; Glycosaminoglycan 9-12; Myristyl 10- 15;19-24; Pkc_Phospho_Site 23-25;
35	DEX0253_246	Ck2_Phospho_Site 3-6;35-38;
	DEX0253_247	Ck2_Phospho_Site 12-15;27-30;
	DEX0253_248	Glycosaminoglycan 30-33; Myristyl 18-23; Pkc_Phospho_Site 22- 24;
	DEX0253_250	Asn_Glycosylation 21-24; Ck2_Phospho_Site 14-17;28-31;
40		Myristyl 26-31;

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have
 45 a phenotype of interest. cDNA is then generated from these RNA samples using
 protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a

template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 147. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*,
5 *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then
10 cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are
15 nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium
20 iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical
25 Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological 30 Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific

-133-

antibodies, at a final concentration of 0.2 to 10 $\mu\text{g/ml}$. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the

5 polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove

10 unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are

15 plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion

20 consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

25 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , $\mu\text{g/kg/day}$ to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day , and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given

30 continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g/kg/hour}$ to about 50 mg/kg/hour , either by 1-4 injections per day or by continuous

subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

5 Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation
10 auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

 The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable
15 polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl
20 acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP
25 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

30 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable

-135-

carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the
5 formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-
10 aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate,
15 succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates
20 including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about
25 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a
30 container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 $\mu\text{g/kg}$ of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the

This page is not part of
the pamphlet!

WO 02-48370

2/3

Date: 20 jun 2002

Destination: Agent

Address:

treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of
5 expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a
subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and
separated into small pieces. Small chunks of the tissue are placed on a wet surface of a
tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned
upside down, closed tight and left at room temperature over night. After 24 hours at room
10 temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of
the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and
streptomycin) is added. The flasks are then incubated at 37°C for approximately one
week.

At this time, fresh media is added and subsequently changed every several days.
15 After an additional two weeks in culture, a monolayer of fibroblasts emerge. The
monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al.,
DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine
sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf
intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using
20 glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified
using PCR primers which correspond to the 5' and 3' end sequences respectively as set
forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer
includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear
25 backbone and the amplified EcoRI and HindIII fragment are added together, in the
presence of T4 DNA ligase. The resulting mixture is maintained under conditions
appropriate for ligation of the two fragments. The ligation mixture is then used to
transform bacteria HB 101, which are then plated onto agar containing kanamycin for the
purpose of confirming that the vector has the gene of interest properly inserted.

30 The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture
to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf
serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then

added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35 (3): 470-479, Chao J et al. (1997) *Pharmacol. Res.* 35 (6): 517-522, Wolff J. A. (1997) *Neuromuscul. Disord.* 7 (5): 314-318, Schwartz B. et al. (1996) *Gene Ther.* 3 (5): 405-411, Tsurumi Y. et al. (1996) *Circulation* 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention
5 may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain
10 sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide
15 production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue.
20 Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to
25 the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin
30 fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 µm cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection

-141-

may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

5 **Example 13: Transgenic Animals**

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific
10 embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection
15 (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et
20 al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723
25 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated
30 oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

- The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail
- 5 tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the
- 10 polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous
- 15 gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.
- 20 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using
- 25 techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.
- 30 Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than

one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to
5 both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited
10 to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

15 Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of
20 the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that
25 contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*).
30 However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including
5 human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence
10 associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the
15 control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the
20 body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

25 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the
30 introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function

-145-

of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

5 All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The
10 present invention is limited only by the claims that follow.

-146-

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes
5 an amino acid sequence of SEQ ID NO: 148 through 250;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID
NO: 1 through 147;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid
molecule of (a) or (b); or
 - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is a cDNA.
- 15 3. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a colon specific nucleic acid
(CSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule according to claim 1
under conditions in which the nucleic acid molecule will selectively hybridize to a colon
specific nucleic acid; and
 - 30 (b) detecting hybridization of the nucleic acid molecule to a CSNA in the
sample, wherein the detection of the hybridization indicates the presence of a CSNA in
the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60%

15 sequence identity to of SEQ ID NO: 148 through 250; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 147.

12. An antibody or fragment thereof that specifically binds to the polypeptide

20 according to claim 11.

13. A method for determining the presence of a colon specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under

25 conditions in which the antibody will selectively bind to the colon specific protein; and

(b) detecting binding of the antibody to a colon specific protein in the sample, wherein the detection of binding indicates the presence of a colon specific protein in the sample.

30

14. A method for diagnosing and monitoring the presence and metastases of colon cancer in a patient, comprising the steps of:

-148-

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the colon specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of colon cancer.

15 16. A kit for detecting a risk of cancer or presence of cancer in a patient, said
10 kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with colon cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein
15 said administration induces an immune response against the colon cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

20

SEQUENCE LISTING

<110> Sun, Yongming
Recipon, Herve
Ghosh, Malavika
Liu, Chenghua
DiaDexus, Inc.

<120> Compositions and Methods Relating to Colon Specific
Genes and Proteins

<130> DEX-0254

<140>

<141>

<150> 60/244,717

<151> 2000-10-31

<160> 250

<170> PatentIn Ver. 2.1

<210> 1

<211> 421

<212> DNA

<213> Homo sapiens

<400> 1

```
cttaaaaata atttctagat tgttggcatt attaaaaccc taaatccttt taggaactat 60
tgcgaagaaa gaatatgata ttcgtaagag ctcaagtcta atattagcat tggttatggt 120
agtgaagac cagataaatc ttttagttgg gaagtatgtc ttgaggata cttccttata 180
atcattaagt aaataagtaa aactatatta catagataat gtgtaactct ctgtattaca 240
tagaatgtct gcagaatgta gataggaaaa ataaagtttg tcaataatth tcaacatctt 300
tattgagata cagttaatct gccatgacga tttgcctact ataaagtgtg catttcagtg 360
tgtttagcta gtgtatttgc agagttgtgc agtcatcacc acagtaactt ccctaacact 420
c
```

<210> 2

<211> 426

<212> DNA

<213> Homo sapiens

<400> 2

```
agaaacccat tcctaagtga actgccactg ctctagtcta acttaggttg gcagagagcc 60
agcactttct tcagcattca gggcagggag cactgaggat attggcattg cttattacta 120
agcacacaga tacaagtatg tgcttgatat gtaaccaaag taagttaaac tccttattta 180
atcttagcac ctgtctaaag gctgggtgac tgtatttata gatgaggaaa actgaaaatt 240
gggggccaag gggcagtgaa gtgaagtgac ttgttctatg atacacagct agtaggaata 300
ttagcactgg aatttgaatt tcatgccatc ccattccaac ctgggtgttt actacttccc 360
actatctccc aagcatgggt attttaggaa atatagaaca ttttctcagc aatacagact 420
tatttc
```

<210> 3

<211> 1016

<212> DNA

<213> Homo sapiens

<220>
<221> unsure
<222> (469) .. (574)
<223> a, c, g or t

<400> 3
agaaacccat tcctaagtga actgccactg ctctagtcta acttaggttg gcagagagcc 60
agcactttct tcagcattca gggcagggag cactgaggat attggcattg cttattacta 120
agcacacaga tacaagtatg tgcttgatat gtaaccaaaag taagttaaac tccttattta 180
atcttagcac ctgtctaaag gctgggtgac tgtatttata gatgaggaaa actgaaaatt 240
ggggggccaag gggcagtga gtgaagtgaac ttgttctatg atacacagct agtaggaata 300
ttagcactgg aatttgaatt tcatgccatc ccattccaac cctgggtgtt tactacttcc 360
cactatctcc caagcatggg tatttttagga aatatagaac attttctcag caatacagac 420
ttatttctct atttccctt ccacatactc tcttttcct taacaacann nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnntcacat catttattct taggccactt 600
tgatgctttt tcattgatgc tctttataga catagtgaag taaaagttaa tctaggatat 660
atgggtgggag gtgaggaaga cttaggtaga gaggttccaa accagttgtt actgcttagc 720
tcaatttcag acatacttcc tccagccctc tctaaactac ccaccagtct tcgccctctc 780
tttcttagtt ctgtggcact tgccctgggt gccctaactg tatggcatgc tgttctcatc 840
agtctgaggtg agactagcat cgaaaggcac atcaacaaga aggagagacg tcggctacag 900
gccaaaggga gagtgagtag ggttgaaggc tcgggggtgg taggtgggta actgaacttg 960
ctctccctgt aaacagaggc catgggcagg gctgactagg gcaagcatta taaaag 1016

<210> 4
<211> 1358
<212> DNA
<213> Homo sapiens

<400> 4
ctcctggggc tcgttttctc caggaggctg cattctgac cataaacctt ctccctgggg 60
tttagggctg agctgttct gatgtttatc ggagactggg atcaaagcta tccagggtcat 120
aaatctctct ctgtggctgt tgggccccag ggcagctgaa gaggggtgac agcccttttg 180
acctcaaagg aaaaaatgtg ctctactcca ccactccca gctctgcaa gaagctgtcc 240
tctgagaagc catggctggg ccgttccatt ctggggagct gctgaaaaga gctgggaggc 300
cgagaagaac ttgcgtgtgc tgggggagag gaagcctggc cttgaggagg ggggtgcagg 360
gtggctcctg tgtgtgtggg ggctggggga ccttgtgtgc cttttccttg tggctgtgaa 420
atgctttatg agtacttcca taggaggatg gacagggagt cggggagata aactcagcca 480
caaggcccca gggcctcagg aaacttgcac ccaaccctct cattttacag aagaaaactg 540
tgctggaag gttgaagggt ttgttcccag tcacacaacc agggatcctt aggacagcca 600
gaccaggaaa ccatttccaa actgccaaag catggcagag tatcaagacc tcaggaacca 660
tcgagacacc atggaagcat tgggaaaagc ctcttagct tttgaagctc ctcatgttc 720
ttgagtgtgc atggagccca tgactgcggg gttttgtaga cacctcaggg attacatgac 780
tggtagccct gacaaagtca aggtgtgtg acaaaatgag tccgaggatt tcaggggcac 840
gctgggcgca ggagctggtg ggctgttggg agtggccctt tactgggcag gcttccctcc 900
tcctgggtgat ggggggttcc tcagcacaaa agtgaagggg tggaggggct ggaggagcag 960
gaatctctct tgttgatagg tatgaggcct tgaagtcctt ttctttgtcc caggattcat 1020
ggacgcttcg gggctgatct ttgagtttcc aagcatgggg tgcagagacg tttaggtaaa 1080
ctcttaccgt cctctctctt cgtcagggtc tcccaggaat caacaatgcc caagaaggaa 1140
gggattgtag aaatagctta accctttcat ttaccaacgt ggaaattgaa gcccagggaa 1200
gggaagggac cggctgtgga agggagagcc atcagcagaa agagaccctg agatcttcgc 1260
ctgggattcc caggaagtcc agcccagct gattcacaga acaaatgcat gcaaaccttg 1320
ctatcaataa attacacatg cacttacgta aaacacat 1358

<210> 5
<211> 2375

<212> DNA

<213> Homo sapiens

<400> 5

```

cttttctctt gttgagtgca aatggagaac agctgctcac gctcgctcgc tgacatcagc 60
tatttctcag gatgacctg cgagacaggc cagggtcatt agaccaatt tggttctcag 120
caaatatgtg tttattcctg catgcgtggg ccacaggctg gtttcttggg tgcaatgaat 180
agctgcaggt ttattagggt gtcttttttag atggatgtat gtttcccgat gtctatagaa 240
cactccggac cccggagagt gaagactctg cctgctggac ttgctttgag aagatccttc 300
tccacctccc catggcagaa gttgcttcac agaggggaac agttttatgg atgtggctga 360
gaccttaaac ttgaggcaac ccactctgagg tggcatccag aggagactgg ctggcccttc 420
cttcaccttg gatgtagtgc tgtttctagg atctcttttc aatcagcaaa acaggggatg 480
ttccaagagg gtgtggatcc cctgccatcc cacatggtca agtggagggg acgggaaaaa 540
gctatgaagg gtttgtgacc acacagactc tcctggcccc ctgtccctttt ggaaagaaga 600
cagggatgaa atataatcaa gcaattaacc acccccatca tcaccaagaa caacagtatc 660
aacaagaaga acagggacaa caaaacccac ggatgaaaca ttcctttctc agctcagatc 720
ttatctgggt cgttctctct ctgctctgtc ttggtgtgtg gtttagagaa acatggacaa 780
cgctgttttg aagaacaggt gagcgagggt ggggaatttc agaggcctgg gccaccgcc 840
tccacctcct cccagttta acctttgaca ggatcttcac ctctctctga tcagcattgc 900
ttcttgttca aaggcctcag ccaccagct gtgtcccttt cccagaaaag caagggcaga 960
tggcagtggg tctgttgatg agagaacttt aagggcccaa tcagtccctg ggcaccctt 1020
ctcgggctcg ttttctccag gaggctgcat tctgatccat aaaccttctc ctcggggttt 1080
agggtcgagc tgttctgat gtttatcgga gactgggatc aaagctatcc aggtcataaa 1140
tctctctctg tggctgttgg gcccagggc agctgaagag ggttgacagc cctttggacc 1200
tcaaagggaa aaatgtgctc tactccaccc actcccagct ctgccaagaa gctgtcctct 1260
gagaagccat ggctgggccc ttccattctg gggagctgct gaaaagagct gggaggccga 1320
gaagaacttg cgtgtgctgg gggagaggaa gcctggcctt gagggagggg tgcagggtgtg 1380
gctcctgtgt gtgtgggggc tgggggacct tgtgtgcctt ttccttgtgg ctgtgaaatg 1440
ctttatgagt acttccatag gaggatggac agggagtccg ggagataaac tcagccacaa 1500
ggccccaggg cctcaggaaa cttgcacca accctctcat ttacagaag aaaactgtgc 1560
ctggaaggtt gaagggtttg ttcccagtc cacaaccagg gatccttagg acagccagac 1620
caggaaacca tttccaaact gccaaagccat ggcagagtat caagacctca ggaaccatcg 1680
agacaccatg gaagcattgg gaaaagcctc cttagctttt gaagctcctc attgttcttg 1740
agtgtgcatg gagcccatga ctgcgggggt ttgtagacac ctgagggatt acatgactgg 1800
taccctctgac aaagtcaagg ctgctggaca aaatgagtc gaggatattca ggggcacgct 1860
gggagcagga gctggtgggc tgttgggagt gccctttac tgggcaggct tccttcctcc 1920
tggtgatggg gggttcctca gcacaaaagt gaaggggtgg aggggctgga ggagcaggaa 1980
tctctcttgt tgataggtat gaggccttga agtccttttc tttgtcccag gattcatgga 2040
cgcttcgggg ctgatcttgg agttttcaag catggggtgc agagacgttt aggtaaactc 2100
ttaccgtcct ctctcttctg cagggtctcc caggaatcaa caatgcccac gaaggaagg 2160
attgtagaaa tagcttaacc ctttcattta ccaacgtgga aattgaagcc caggaagg 2220
aagggaccgg tcgtggaagg gagagccatc agcagaaaga gaccctgaga tcttcgctg 2280
ggattcccag gaagtcagc ccgagctgat tcacagaaca aatgcatgca aaccttgcta 2340
tcaataaatt acacatgcac ttacgtaaaa cacat 2375

```

<210> 6

<211> 410

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (34)

<223> a, c, g or t

<220>

<221> unsure

<222> (56)

<223> a, c, g or t

<220>

<221> unsure

<222> (108)

<223> a, c, g or t

<400> 6

```
cagagtc aag gccc aaggc cgtgggtctt tggaggagg gtttttgaga catgtncagg 60
gacaaaccta gcaacaagag aactcttaaat cccatacgtg atattgcnaa ttagcttttc 120
ctttcacaaa tattgtccac cctaagtatg tttactataa tgttagctgt taaagacccc 180
tctaccccc aaaccattta cccttcaata aaaatggtgc caagttgcaa gggtttagaca 240
ggtatgtatt gaaatttaga aagtttgaat aatttcttta acacaaaagc atttttttct 300
tatttctcat acttttgaat ctattttaat acaacttcag tgctgattaa tctactaaat 360
gtgaaagttt aagatttata gctgggtgca gtggctacac ctgtaatcct 410
```

<210> 7

<211> 416

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (114)

<223> a, c, g or t

<400> 7

```
ctcagagcaga gtcaaggccc caaggccgtg ggtctttgaa ggagggggtt ttgagacatg 60
tacagggcaca aacctagcaa caagagaact cttaatccca tacgtgatat tgcnaattag 120
cttttctttt cacaatatatt gtccacccta agtatgttta ctataatggt agctgttaaa 180
gacccctcct acccccaaac catttaccct tcaataaaaa tgggtgccaag ttgcaagggt 240
tagacaggta tgtattgaaa tttagaaagt ttgaataatt tctttaacac aaaagcattt 300
ttttcttatt tctcatactt ttgaatctat ttaaatataa cttcagtgct gattaatcta 360
ctaaatgtga aagttaaga tttatagctg ggtgcagtgg ctacacctgt aatcct 416
```

<210> 8

<211> 786

<212> DNA

<213> Homo sapiens

<400> 8

```
atgttcctag tagaacacaa agtttgctca ggtaacacac aagtaagcat taaatgcctt 60
cctgttgat ctgagaagtt tgttatgaaa tattttggta accgctgcat agtcagtgtg 120
ggaggagcag atgaatttta gctgtggtta tgtgtgctgt aaaagactat acgtgcttgt 180
attagtcaga atgagtacac cactaatttt tgtatggtta gagatttata ctaagctcat 240
catcagtttc tataattcag tgagataaaa ctgagtcaga ttgattttta ggtagcacat 300
gtagaacacag ctaattttat tcccctgatt tgatcctcat ctattgatta tataaactaa 360
agaagctaag aacaattaac ccttacgagg ttacacagtc aggagatgct gaactgagat 420
tcagtgtaga aagtctgtct tcagagccta tgcttttagt ctttatgcta agtttaactt 480
gtttaaatag caagattatg aagcactata cagtgcctc gtatagacaa aaatatagta 540
tattgattat tagagaaact acatattaga ctgttgtaga tacgtgggca agtatttggt 600
aaatcatttc agttgcctaa atttaagcaa ctgtgctgtt taaaacatgc tcattcacat 660
ttttctttaa tctagaaagt cactctgaa taattgcttg tttagatttt ctcatttggt 720
gtgggaaatt tatattaaaa ttttaactaa tattctaaca atacagagtc tgaacctaaa 780
gtccac 786
```


<210> 9
 <211> 1509
 <212> DNA
 <213> Homo sapiens

<400> 9
 atcagaccta gtgcgtaggc ttctggatct cagaatcact tataacttaag tccaggetgt 60
 tctcaaataa ggcaagaagc atctgctgtt aatagctgac agtaaattac acaaagtaaa 120
 acatggaaaa ttaaagtcag aaaagctagg aagcttttct atcattttca attttctgca 180
 aaaatacaga cataatcagg tttaggatct gcttgtgatg gataaattac atctgtaatt 240
 ccttcttttc catattactg cattcagacg ataatttgct ttcagatatac ttgctcatct 300
 aatcgttcat agactggaaa taagtagtaa catctcccaa tcctaggaag catttataac 360
 tagtctttgc ctttttgggt gttgatagac tagtggtgat tataagcttt cgagcttctg 420
 aaaagcacia cgaagattaa aataatcata ggataataaa atacttttaa acccttctag 480
 tctttaattt taaaatgttc cagtagaaca caaatttgct caggtaacac acaagtaagc 540
 attaaatgcc ttctgtgtga tctgagaagt ttgttatgaa atattttgga aaccgctgca 600
 tagtcagtggt aggaggagca gatgaatttt agctgtggtt atgtgtgctg taaaagacta 660
 tagtgcttg tattagtcag aatgagtaca ccactaattt ttgtatggta agagatttat 720
 actaagctca tcatcagttt ctataattca gtgagataaa actgagtcag attgattttt 780
 aggtagcaca tgtagaaaca gctaatttta ttccctgat ttgatcctca tctattgatt 840
 atataaacta aagaagctaa gaacaattaa cccttacgag gttacacagt caggagatgc 900
 tgaactgaga ttcagtgtag aaagtctgtc ttcagagcct atgcttttag tctttatgct 960
 aagtttaact tgtttaata gcaagattat gaagcactat acagtgacct cgtatagaca 1020
 aaaatatagt atattgatta ttagagaaac tacatattag actgtgtgac atacgtgggc 1080
 aagtatttgc taaatcattt cagttgccta aatttaagca actgtgctgt ttaaaacatg 1140
 ctcatcaca ttttttctta atctagaaag tcacttctga ataattgctt gtttagattt 1200
 tctcatttgg tgtgggaaat ttatattaaa attttaacta atattctaac aatacagagt 1260
 ctgaacctaa agtccagaag aattttaagt catgccgcag acaggatgaa cagtatagca 1320
 aatcagaata atagactgtg agggggggta ggggggaacc catgagaatt tcaggatgtc 1380
 aagataaagc ttggaattga ggtaaaggca tcagataagg aagtgatcat ttcataactt 1440
 gtttttgctt gaaatatatt atattttaca tcacaaaagt agtataactg ttattttgct 1500
 aatgcacag 1509

<210> 10
 <211> 283
 <212> DNA
 <213> Homo sapiens

<400> 10
 ctaagtaatc cttgtcaggg gaggtgggtc ccaattcgtg actcttgac cttggggcat 60
 cttatgattt attgttatca ctaacaatag ctcgctatgt gtcagtgtct ctgctacata 120
 ttttatgttt tatttcagct tttaaaaaga ttttcatgat tcatgattgt tgtaaagcag 180
 gactaggctg tatgtacata tttgaaatga aagtttcaca aaacatcatt tacctttact 240
 atgtgtgaca cactttgcta ttttccattt aatctatttt att 283

<210> 11
 <211> 736
 <212> DNA
 <213> Homo sapiens

<400> 11
 gtctttctga aaggaagcac tcggaatcct tccgaacttt ccaagtccat ccatgattca 60
 gagatactgc cttctctctc tctgggattt tatgtgttct tgatagttaa ttgttgatgt 120
 atttgctact ttgcttcttt tctctttcaa gacttgatca ttttatatgc tgtttgaga 180
 aaaaaagaac ttttgtagc aaggaggttt cagaaatgat tttggatttt ctgtaagtgt 240
 ttaatttagt tctaggggac agcatctctc atcccggagt aaatttctgc ctttgacctg 300
 catggattat tttttcaggc tgcggaattt ctcggcacct acctgtagta tggggcactt 360

6

```

ggtttggttg cagagtaaga aggtggaaga atgagctgta cttgggtaag cagttgaaac 420
cttttttgag caggatctgt aaaagcataa ttgaatttgt ttcacccccg tggattccag 480
tgggccccgac agcgcaacag gtttgcagat ttcttttgaa attccttttt cccccctccc 540
tctgcctcag caaaagaaaa gaatccatat aacagggttca tgttcaattg cttgggctttt 600
cagcacttat tctgaagact ttataatatt tttaaacttg accttggaac acagagggct 660
ttgtgggtga ggtgtattta tatttactta aggggtgcaca ttttaaaaat cttattctgt 720
gtttgtacaa agacgc

```

736

<210> 12
 <211> 547
 <212> DNA
 <213> Homo sapiens

```

<400> 12
ccggtttagaa tagagcttcc acaagctcct actttgatat ctgccctcct agcactgggg 60
ccactgtttc ctgctttccc tctatgtgaa ctctccgtgt ttctaataac atctggatta 120
atcacatcct ctctggccta ctcaaagata gtaactctaa caacttttcc ctctctttca 180
tgcaattcct actttgcctc tctctgctgg actttttctc atcgacatat aaacatgctg 240
ttatgtctcc caacaaaaaa aaatgcaaaa accctttcag ccctatgctc acccatcatc 300
cagctgtagt cctcttccct ccttttactc tcctttatta tagctaaatt tcttgaaagg 360
atggaatgtc cacttcctct cctcccatcc ttctctgaac ctaccccaat ctgccctttg 420
tccccactgt gccagtgaga gggctcttga taagctctcc cttcattgac ttccagttgc 480
tcaatgaaat gggcagttct cagtcctcat cttacttgac tttccagcag catttagtac 540
taccaga

```

547

<210> 13
 <211> 1559
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (1337)
 <223> a, c, g or t

```

<400> 13
gttctacgct taaaacaacc tcttcccct aactttaaaa tcagatacag taaaagcctc 60
ttgttgagga tgtggttata ttggtagatg agagtgtgtc agaaacaggt agaaacttac 120
ctagcaaaag aactagtact gtatcttgac ttgttacatg gcaacaatca attagatgat 180
aatttctatt taaaagcatt ctatatgggg aaagacatgt tcattttgat aagtaaagac 240
aaaatctagg tttttagtgt atgtgtgttg tacatgtggg ctttggaag caaacctaac 300
tatgtattat tgacattaaa aatgatgact taatgctggg taaatcctgt actcagaaga 360
tactcactga tgatccattc ctggctataa cctatgaact aaacgaattt tttaatcttg 420
gtgcttatta ttagcttcag cttgcctctc taataatccc aacaccttgt gctctcatcc 480
tgctctcagc ttattacttt gccccgtttt tcaactgagaa gacagaagca gttagaatag 540
agcttccaca agctcctact ttgatatactg ccctcctagc actggggcca ctgtttcctg 600
ctttccctct atgtgaactc tccgtgtttc taatatcacc tggattaatc acatcctctc 660
tggcctactc aaagatagta actctaacaa cttttccctc tctttcatgc aattcctact 720
ttgcctctct ctgctggact ttttctcatc gacataaaa catgctgtta tgtctcccaa 780
ccaaaaaaa tgcaaaaacc ctttcagccc tatgctcacc catcatccag ctgtagtcc 840
cttccttccct tttactctcc tttattatag ctaaaattct tgaaaggatg gaatgtccac 900
ttcctctcct cccatccttt cctgaacctc ccccaatctg ccttttgtcc ccactgtgcc 960
agtgaagagg ctcttgataa gctctccctt cattgacttc cagttgtctc atgaaatggg 1020
cagttctcag tctcatctt acttgacttt ccagcagcat ttagtactac cagccagtcc 1080
tcatccttga aatactttct tttcccatat ctctaactgc ttaagtcaaa agggttccat 1140
gatccagtc ttacataact taccttcttt ggctacgctc attatctggg atctcatcca 1200
gtcttggggc tttaaatact atatggggac aactacagcc gagaaccttt ccctgaactt 1260

```

tagactcttt tgtccagaag attatacaaa ttctctgttt ggttatagaa tttagaatgc 1320
cccaaatcaa gataatnctc cctcaattct gttcctccta taagcttccc caatcggtaa 1380
atgaaaactg tgtccttcta gttaatcata ccaaaatcct aaaaatcatc cttaactcct 1440
ctcatctctg atatccatat ccaacccatg agcaaatact gtcaatctgc cagaatccaa 1500
acatctctcc agccccattg ccaccaccct ggtccaagcc accaccaggc cttgcctag 1559

<210> 14

<211> 1455

<212> DNA

<213> Homo sapiens

<400> 14

ggagtgtgaa ggtggtgagt catgggagtt ccaagggaaat ggggtataaa gggaggtctc 60
aaatgaggca caagtggaga aggtagcttg ggaaaggaga aggatgcttc tccttataag 120
atgggaaagg cagaggaaga gggccaagat acagtgatct aggggtgata tggaaagtga 180
ttgagagaac tcaactctgg gttctgaaac ccctaggttt ggggggcttt gagataggga 240
agagggttaa agtcagttgt tctagcaaat atgggttgga atttatttgt gatgcttaaa 300
aatattgctg aagagaagtg aagtctatcc tagagttgga tgggtgagatt atttagtga 360
actaccagat ccatgttggt attccttcca gtatcattca gcagcccttg ggcagttgcg 420
aggcaagtca tcagtgggtg atggagattt tcccaggtgg gtgtggttga aggcaggga 480
gaacgagttc aggagcacat tacaagaaga aggtgactgt aagggtccagg ctgagcagga 540
aggtaaaagca agaaggaaac atgaggttgt gaagagaagt ttagagggat gaggaggcag 600
gagagatgaa cagttgcagg atgtagctag agtggcgatg ttagatcttg gggccagaga 660
tctttacaat gattatgaag atcaaagggc attagaatca agctataaag agccactgtt 720
tgatgttggg atgtgaggat gctgcagggt gatgtctgca cattgatggt gagaacatgg 780
tcacctctgg cctgctgggt ctttgctaaa gagactgtgc tctgttcttg gggccgtttt 840
catcatctga ttagagcagt ggtccccaca tgggtgttct tggaccatct gtataaaatg 900
ttcataggtc aaggataaaa tggaaaaaca gagaaaatgt cacagaaatg tgcccattgt 960
tgaaagacca ccagctgtcc tttttggagg attgttcttt attctaaaaa tgtatatatt 1020
ctattctatt aaaacatttt tgtattggca ttttttctc ttttatgaaa tgccatgggg 1080
tagaaatttg taatgtatcc aattctctg tcttcatgta ttgccctgtg gtgggggagg 1140
ggagtgtggt agtactggcc aagaggctgg gggcagaggt gcaatgttag acttctagcc 1200
tggagcattt aattcttagt acaagactct ctaacattct tctccctctg ttccctgctt 1260
ggtgatactc gaggtattgc aacccccatt aaccttagtc ttagggcaag tttgatggga 1320
aacagagcac cccacacctc cctgcagatg aagcatgagt gagaaaaaca acttctgatg 1380
tttgaaagta ccaagatttg ggagttgttt gttattgcag caaaacctca cctattctga 1440
ccaatcatgg tggaa 1455

<210> 15

<211> 904

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (281)

<223> a, c, g or t

<220>

<221> unsure

<222> (329)

<223> a, c, g or t

<220>

<221> unsure

<222> (469)

<223> a, c, g or t

<220>
 <221> unsure
 <222> (471)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (539) .. (540)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (662)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (692) .. (693)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (701) .. (703)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (776)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (785)
 <223> a, c, g or t

<400> 15
 tggaaaacca aaaaattgat atgctaaagt atactcttaa aggtcttaat actttaaaag 60
 tatatagatc tcatgaacat aattcatttg aggaaaaaaa tacaaatcat ttcttgtccc 120
 aggaaaacag taaatcttta atggaacttt ttagcaatta tgacaaaaag aatggaaaaa 180
 tgtttaaaaca tatataaaaag gctagacgtt tategccaaa tagtatctaa aggtcataga 240
 atagtttaga attctgtcat ttgttttgt gtaataaata nccccctcct ttaccctttc 300
 accctaataa tagatatcca ccattttgnt gtgattatcc aactatagag tacctttttc 360
 aagaactcat tatataccaa agtaggagct tgctgacact gataatgctt tatttagttt 420
 tgtagtgaca tacaattacc atttgcttag gaaaaaaaat aaagaacana nacaagtaaa 480
 ttttttaaaa ctatggttgt gtatatataa gttgataaaa atcctttggg agaaaactnn 540
 tgtcttgtgt gttaagagca ttaaatagtc ataccctta gcctagtgtg tcttctatcc 600
 tgaaaaaaa ttaacaaagc aaatactaac ttaagaaaaa aaactacagc actgaaaaga 660
 tntgttgtta tattgtttat gctaacataa annatgtaaa nnnttatata ttgtttatac 720
 tgacttataa tttattacta tacatagtgt aaattatgat acattggcct tggtangcag 780
 ttttntaacc gctaataata taaataccat actattaaca atctagaaaa atgattctgg 840
 tataggttat gtgaaaaggc acaaaataaa attgtatata gtacactagc aatgaacagt 900
 ctga 904

<210> 16
 <211> 984
 <212> DNA
 <213> Homo sapiens

<400> 16

```

acagattttac tctcctgaat tttccagaaa tgtagatact tttaaatcaa aggaaggctg 60
tattttgttt tggttcagaac ttttctattc cagaaaatca tgtcaattga cagcaaagcc 120
acttggtggc attgagcctc ctgtgtaaag caccgacgtc attctgtagt tgatcatcact 180
gtatttcaggg tgattctaca cgtaggagtg agcatttgac agcttccatg tcttctagt 240
cggctgagaa tttacatatt aagatacaca ttatttatta tcaattactt tctgtttca 300
atgtccatctt agagcactaa aaatatcttt gtaggtagtt gatattactt atgaatttta 360
tttcaggaga gcaaaggaaa atacaagata gttgtatgaa aagggggcac cgggtgtgct 420
agagtggctc accaccgccc tacacagtgg gctaattggc tggagagtag agctgactct 480
gcacagttgc atgctgaccc tctgaagaat ttttttacia aagcgtgacg tcgctggaag 540
accttgacag aattagcaaa gcggttgaga tgcatacttt ggagtcagac agactccagt 600
tcacatcttg gcttttatac ttacagctgt ataaccgtag acaatctatc taccctctgg 660
ccgactccat ttctcattt ataagatagg ataactgtg aaatgctttc cacaagatta 720
ctattgcatt tatttctctc accactctta atgaagagag tcttgtaaca gataactcta 780
attgtcttca gagttcaggt ccccaagaaa gattatgcct tctaaaagct agtctgtttc 840
cttcagtgag gagccatttc attcatgctg ctctactctt tacttgagct gctagcaaac 900
atggagctaa gtactcatgc ttaatttctg tggctttcct caaatagggt ttcaatacta 960
tagtttgccc tcaactccatt ccct

```

984

<210> 17

<211> 429

<212> DNA

<213> Homo sapiens

<400> 17

```

cgtgataaaa atagtttgct ctgagttttt gcctttctgg aatttaatat caagaaaaat 60
atgttcccta ccctctcagc cccactcta cctccctgtg gcttgtaaag ccttccttct 120
gcctcctgca tcaacttctt gatggagagt gtatgaatgc aaaagctcct cccttagcac 180
ttacctagt ctctactctc tgggtcctg ccactgggtc ccagctaaga gagtttgatt 240
ttaaaatcca gagtttatgg ctttttaaaa ataacctctc acctatttat caaaagctcc 300
ttctaaataa tatttacaac aacaacaatg ataatggcta ctatctagta ttcccatctt 360
tccagacact gtgctgggct ctttccaaac actgttttaa tctttaccaa caccagctcc 420
gccgctcta

```

429

<210> 18

<211> 734

<212> DNA

<213> Homo sapiens

<400> 18

```

cttttgagacc ataagcctca ggaagctata aggattatct gcattcttac acctgggcac 60
tcttcccttt tgctgaatac cagtttttca atcttttcta tttttgaaat aggttaagaaa 120
agaaaataat tttctagaat ttgaagaaaa atcttaaaac atttgaaatt ctttgttatg 180
atgactaata taacgaatag cactcagggt tatcaaata taacattttt ccatatttgt 240
tatagaatct ttttccatat ttgctacaga aataatttct ttatatatat aatacatatt 300
tgaacactga ttttacttga tacattaata taatgctgat gtgctgagat gaataaatca 360
aagaacctct tggagctctt ggtgtgcaat aagcatagtt aacgaatata aaataagtga 420
tattttctag aaaataaata ctggtctaca atgccttctc tgcattttca aagctctctaa 480
aaagatctga aaatccaatg ctttttaaaa ataaaattac ggtaactctca tttggccaca 540
aaacctgttc agaattgatg tgaggctatt aagatattta tttctcttat ttattagtga 600
atattcatct ttactacag aaatactaac gagtttgatt acagggtgct ttagacttcc 660
ctcaaggtgt acatatttgc tacttttctc taaaatccca aacatcctgg attctgaaac 720
acatctaaac cccc

```

734

<210> 19

<211> 1184

<212> DNA

<213> Homo sapiens

<400> 19

```

attctaactc tgtgacatgc agtctgtgac actgagagtt acttgacact tctcttgga 60
tgagatcctt ttctagtgc gacattttat aattctattc tgtatcgtgt tcatttaagt 120
agtctgcttt atcattacat taacattttat gaaagacttg ctggtatcat tggcttagcg 180
attatttttc catctagatg ctttttttaa agaaatgaag agaataatga atgttttaaa 240
tgtacatttt agtttgattt aaattttaat caaggatttt tattttatac attacatact 300
gatcactggt ttatgttaac tctggtccta ataaacagaa aataacaatt tgggaatatc 360
acaacaatga gagctcgagg taaaatatag cataaataag acatatatgt gtatgaactg 420
agatatatag aaataattaa atgtaacaat cttttggacc ataagcctca ggaagctata 480
aggattattt gcattcttac acctgggcac tcttcctttt tgctgaatac cagtttttca 540
atcttttcta ttttgaaat aggttaagaaa agaaaataat tttctagaat ttgaagaaaa 600
atcttaaaac atttgaaatt ctttgttatg atgactaata taacgaatag cactcagggt 660
tatcaaatat taacattttt ccatatttgt tatagaattt tttccatat ttgctacaga 720
aataatttct ttatatatat aatacatatt tgaacactga ttttacttga tacattaata 780
taatgctgat gtgctgagat gaataaatca aagaacctct tggagctctt ggtgtgcaat 840
aagcatagtt aacgaatata aaataagtga tattttctag aaaataaata ctggtctaca 900
atgccttata tgtcatttca aagtctctaa aaagatctga aaatccaatg ccttttaaaa 960
ataaaattac ggtaattctca tttggccaca aaacctgttc agaattgatg tgaggctatt 1020
aagatattta tttctcttat ttattagtga atattcatct ttactacag aaatactaac 1080
gagtttgatt acaggggtgct ttagacttcc ctcaagggtg acatatttgc tacttttctc 1140
taaaatccca aacatcctgg attctgaaac acatctaaac cccc 1184

```

<210> 20

<211> 550

<212> DNA

<213> Homo sapiens

<400> 20

```

ctttcccgct cccggcccca gtgccttgca tgcagcaagg tcttggcatg tgcaagcttc 60
cttaaggagc ctgcagcttt gctccaaagc acacactggc agaccttggc cagatgcctg 120
gcacaggggc tggggaggga aaggctgcc aacccccgtt ttccctttgc agatgagcat 180
tctccaaatc catgtttacc cagtctcct taatgctgcc ttccaaactg tcagcgggtg 240
ctaaaaagca cacattagga tgaattagaa catgccaggc tgcaagggcg ggtgtcatcc 300
cagaactcac agagcacggt gagggctcag ccgctcagcc acatctttag gtcccaccag 360
catctccccc caggcatgga cctccccaat ttaccctgtg aaggctgcat ggagaagatg 420
caggtcttag gaacagccag catcaccaga ggtgccactt agtgagtacc cagtgggctc 480
ccaacaccgt gctgagctcc cagtgggaga accggaaccg tctgcctgtt ctctgttgta 540
ttccagcatc 550

```

<210> 21

<211> 599

<212> DNA

<213> Homo sapiens

<400> 21

```

tactatgtgc cagacacagg agttttcagg atgagtcaat aagataataa acacaaagtc 60
ccggccccag tgccttgcat gcagcaaggt cttggcatgt gcaagcttcc ttaaggagcc 120
tgacagctttg ctccaaagca cacactggca gaccttggcc agatgcctgg cacaggggct 180
ggggaggggaa aggtgcccc acccccgtt tccctttgca gatgagcatt ctccaaatcc 240
atgtttaccc agtctcctt aatgctgcct tccaaactgt cagcgggtgc taaaaagcac 300
acattagat gaattagaac atgccaggct gcaagggcgg gtgtcatccc agaactcaca 360
gagcacgttg agggctcagc cgctcagcca catctttagg tcccaccagc atctccccc 420
aggcatggac ctccccaatt taccctgtga aggtgcagtg gagaagatgc aggtcttagg 480

```

11

aacagccagc atcaccagag gtgccactta gtgagtaccc agtgggctcc caacaccgtg 540
 ctgagctccc agtgggagaa ccggaaccgt ctgcctgttc tctgttgtat tccagcatc 599

<210> 22
 <211> 618
 <212> DNA
 <213> Homo sapiens

<400> 22
 gaaaaactac tctttttggt gtaaagatat tttttatatt ttctttgctt gtaaagagtt 60
 attatcaatt tgtaagtata aaaactgcaa gtatagtgg tagttgataa gaaaggtaga 120
 taataaaact taaaagggat ggacacagat tgaaaaaggc cttgagtgcc aagacaagag 180
 ctctgaactt taacaggcac tggaaaccgt cataggtctt aggtaggaat atgctgtgct 240
 ccaccatct taattaggtc ttatggaggt ttgatagcaa gagggtagga atatcattta 300
 gcaggctact gcaagtatcc aggtgaaatg tacagaggtt ttgaactagg ctgctgggga 360
 ggggtgcagag aagaaatatt ttggaaataa aatggacaga aagtgtataa atggataaag 420
 agaggaatag aactgacacc aggcttcaag cctgatgcct gagaataaag gtgtaattat 480
 gaagggaatc caggaagaca tggaaagagt gggtggagta aggttaaagt gatagtttta 540
 gattgggtta ttttgacgtt gaagtgttga ccaacttctt aagtgaaaat gtgcaacagt 600
 cattgaaaat atgagttt 618

<210> 23
 <211> 711
 <212> DNA
 <213> Homo sapiens

<400> 23
 gaaaaataag tttttgttaa tggttgggat tttcttactg gcctcgtggc aagttttgtt 60
 atctcttatt atatatattc taccttttta tgggaaaaac tactcttttt ggtgtaaaga 120
 tattttttat attttctttg cttgtaaaga gttattatca atttgtaagt ataaaaactg 180
 caagtatagt tggtagttga taagaaagggt agataataaa acttaaaagg gatggacaca 240
 gattgaaaaa ggccttgagt gccaaagaca gagctctgaa ctttaacagg cactggaaac 300
 cgtcataggt cttaggtagg aatatgctgt gctcccacca tcttaattag gtcttatgga 360
 ggtttgatag caagagggtta ggaatatcat ttagcaggct actgcaagta tccaggtgaa 420
 atgtacagag gttttgaact aggtgctgtg ggagggtgca gagaagaaat attttggaaa 480
 taaaatggac agaaagtgt taaatggata aagagaggaa tagaactgac accaggcttc 540
 aagcctgatg cctgagaata aagtgtaat tatgaaggga atccaggaag acatggaaaag 600
 agtggttgga gtaaggttaa agtgatagtt ttagattggg ttattttgac gttgaagtgt 660
 tgaccaactt ctttaagtga aatgtgcaac agtcattgaa aatatgagtt t 711

<210> 24
 <211> 547
 <212> DNA
 <213> Homo sapiens

<400> 24
 aacaaggtaa gcatagccgg ttttcatggg cttattttct catggaaatg attctgtgta 60
 gaattgatta ttcatgaaga cacaatgtaa catcaagttt gggttaatgt tcctcagtc 120
 aacaacaaag acgtatttgt aatcactccc atgagctctac tttgcagcaa gaacatgcat 180
 tttggaatta tccccatcct gtgtgctgaa tactggatgt gactcttagt cagctctgtg 240
 acccttgtca agtaacttaa gctctttgat catcagcttt gtcactctgta aaatgggcat 300
 tctgcctact tcaaagagaa gttgaaggga ttaaacgaga taacctacaa agagcaccca 360
 gcacaatggc ctaaaaaagg aaggcactga atcattctca ctcccctacc ttcagtctga 420
 tctgtctctt attgtcaaaa ggataatttc aattttaata gatctgagat cctgtttttt 480
 aataataatt ttatagaatt tttcatttta tggcaggcac agggctcatg cctgtaatcc 540
 cagcact 547

<210> 25
 <211> 549
 <212> DNA
 <213> Homo sapiens

<400> 25
 gcaaagacct catgaggggt caacgagggg aagccctcgt gggtcagagt acgccacggg 60
 acagactatg ctggcagctt ctagatcggt gaactctggt cttgaagact gggcagaatc 120
 taggaagaac ggaggcacct gagttcacca ggtgggacga acctggcctt agcacggaat 180
 gtggcattta ggtgcttaag tttgttggtt tttttaaatt aaagtgggtg acctggagag 240
 ctgggtgtgga aatgtagcag gaggtctatt tggaaagaag gatggagtag attatgaaag 300
 ttcttaaata tcataatgag gcttgtggat tttattctgt ggtttggatg ctctcttctt 360
 ccattcccttg gatgccaaca ggcattgcact gtttaatctt ggaattcaaa cgggtggcctc 420
 aaacagttag gctgagtagt tggcctcatt agcttcagac ccagcagggc tgggctcaca 480
 ggcgtgtcat ttatcaaggg cttgaatctc tgccagctaa tttatctaag acaactctat 540
 gagatgggg 549

<210> 26
 <211> 350
 <212> DNA
 <213> Homo sapiens

<400> 26
 ctttaagata gatgggtaca catattatga atatactttc cttttgccag accttgacat 60
 tctgtagact tttaatggaa tattatttgc ctctttcatc ttaccttgac gtatgaggtg 120
 gatggccttac gtgcagggtg atgtatgaac cttcccaagc tctgtacaaa tataacttgt 180
 cattcgtaga gacgtatgta tttatatgtg tgcattgcagt cttatttcta gattttcttc 240
 ccatttgctt aatactgaac gctatggcct agatgtgaaa tttaccaggt actactcata 300
 gcaggcagtg aaaccgtgga ctcagctgct ctttccttct ttctctccca 350

<210> 27
 <211> 627
 <212> DNA
 <213> Homo sapiens

<400> 27
 ccacgcgtcc ggtttcaaaa aagaagagta agtcaaagg taaacttttg gggcggagga 60
 aaaaggataa gaaagaggat acagagttta atcagagttg gcatcagata gagtaaccat 120
 ggacatttgg aagctgtaac ctctctcata tttcgccaag gataactgct tctgtatta 180
 tcatgtaatg agttttatgc gtgatggaaa atgtaaaagt aatcttaacc caaacctgca 240
 ttttaatgcc acatggaccg gctgtaattt atggcatctt taagatagat gggtaacat 300
 attatgaata tactttcctt ttgccagacc ttgacattct gtagactttt aatggaatat 360
 tattttgcctc tttcatctta ccttgacgta tgaggtggat ggcttacgtg cagggtaatg 420
 tatgaacctt cccaagctct gtacaaatat aacttgctat tcgtagagac gtatgtattt 480
 atatgtgtgc atgcagtctt atttgtagat tttcttccca tttgcttaat actgaacgct 540
 atggcctaga tgtgaaattt accaggtact actcatagca ggcagtgaat ccgtggactc 600
 agctgctctt tccttctttc ctcccca 627

<210> 28
 <211> 548
 <212> DNA
 <213> Homo sapiens

<220>

<221> unsure
 <222> (132)..(348)
 <223> a, c, g or t

<400> 28
 gttgcatgtg ttggggatat ttctccatta gcaagaagtt tccaaacctt accagtgttt 60
 tgatgaatct aggaacagat ctggcagtga gacctacatc cattttcccc acggacagca 120
 tcttgctgga gtnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnntc ctcctgttaa 360
 atgttgatag aaaaatgggt tgatggcagc atatatccag attgtagatt tcataatatt 420
 aaaggggagt gggcaaataa taaaatgcaa gaaatgaaag catttgaaaa tttagaggac 480
 agaaatgact ttttaagtaag tgatttttagg tgtactggaa tgagtaatct agaataattg 540
 atatgaga 548

<210> 29
 <211> 988
 <212> DNA
 <213> Homo sapiens

<400> 29
 aaatccacaa ataataatth acatttgaga aaatccccc gtactttctat gaataagatc 60
 aagggcaaaa gtgtgctctt ttacatgcca gaaacctcaa gaatttttctg taaggtagac 120
 ttcaaggaaa accaagcagc tcttgactca acaaataaaa atgtaagtct gctgaagaa 180
 ttagtgaacc agggcaccga gtcagctttc tcctaaaaata aatttgagga gctgaaagat 240
 atggatgagg tcagatttct aaaaaatcag tatacacaca gtgttttaag aataaaaaac 300
 agattgatta aagggaaaaa taatttgtaa ataacagaag ccataactta gagataaaaa 360
 taactgtcct ctgattaaca gaacttttag aatgatgaga aaaattaata acacagttaa 420
 agatatcaca gtgattttta aaaaatattc aaggttgaag aaaaaatatt cctatgagaa 480
 tacaggctga aaaagatcaa agtaaaatga atcaggctcg tatcagaaat ttcagtgata 540
 tacaatgaag gaataaaatg gagcagcagc tatagttttg aaacaaaatg tattttccaa 600
 ggttcttgta cccaaccaa ttataactta tgtgttagga caatagagaa gtaatttttag 660
 ccaaagaaat aatctgaaat tatagcattc atgcacattt attgaaacaa gaaactcaga 720
 aatcaaaaata gccgagaaat taataaaaata ttcaaaagga ggaaaataca ttttagaata 780
 aagcataatg aggaataaaa tcactatgac tttttgaaag tataaaaatt gttatttttt 840
 tctatgaata cttgctcaaa tttaaagtag tggattttaat gttgtagcgc taagtattca 900
 gccaaagagt agaactaata aataaaaatg atagttcttt taaaaaaaca taaaaataat 960
 tatctcatga gtagcctaag aaaaaagc 988

<210> 30
 <211> 651
 <212> DNA
 <213> Homo sapiens

<400> 30
 acaccaaata aggtaatgga gataaacttt agaaatcatg tttttaaact gatgtttaaa 60
 agggatggaa ctacacttat ttaaaagggt aagactgcca cgtcagtgtg aaattgttta 120
 aaaaagtcca acacatttggt ggctggacac accagtcaaa tgggtgaaat tagaagatgg 180
 ggaaaaataa tgtcaggtaa atactttatt tcattggatt tatgacttcc cctgtaagaa 240
 gcattattat tttatataaa tacccaaaaa aaaaaaacaa caaaggcagc taattctga 300
 aattaattgc atatgcatca tgatttcaga tatattaac tgtgaaaaaa gtgcgttaaa 360
 atggtaaagc acaataatca aaataaagtt tgtatagcaa tattaatatc acataaaaata 420
 taaattagaa caaaaaagca cttataggga taaagagaaa caccagagaa aaacaaagaa 480
 aaaatcctaa gaaaatataa ctttcacata cttatatggt ttaacagcaa agcccgtaa 540
 ctgtttaata taggaagcac aaacgtgact gaagttacaa gagactgaga caactttcaa 600
 aactcatggg gggagaatth tatcacttca acagaaactt aacaatttaa c 651

<210> 31
 <211> 553
 <212> DNA
 <213> Homo sapiens

<400> 31
 actggacttc ctctttcttc catcaaagac taagatgcct ttttctcttg atgtacttta 60
 ttttgtggag catattatct actttcctga aaaaatgggt tatgggagat aaatcataaa 120
 aaggttttat tagattctac atctcatgat tgatccaaaa gacgttttaa aaacaaaaca 180
 aaaaaaggcc ttgtaggtct taactcttac ttagcctcac atttatttga tagtttgagt 240
 gagtatctta aaaattgaag atgattataa aaattttaat gtagacatta ttttttctca 300
 gaattttgaa ggcactgctc tgtcttttgc agttggagag tctgatgccca ttctgattct 360
 taaatctttt atacaaaaca tgttttttgc tttggcagga agctttacct tttctttctt 420
 tcaagtgtcc tgaacttcca ctgagatgta tcatgggata ggtccacttt gatccactgt 480
 cctggacact tgctaggcct tttcagtcct gaagctcatg actttcaggt aagagaaatt 540
 tacgtctaag acc 553

<210> 32
 <211> 2159
 <212> DNA
 <213> Homo sapiens

<400> 32
 ggccgcttaa ttaaagatct tttttttttt ttttttttag tgctgaataa tagtccattg 60
 tctttatgta ccacagttta tccactcacc tactgaagga catcttagtt gcttcaatgt 120
 tttggagggt acagataatg ctactataaa catccatgtg caggtttttg tgtgaatgta 180
 aagtttccaa ttcatttgag taaataccaa agcatgcaat tgctacatca tataaaagta 240
 tgtttgttac tataaagaac tgccaaactg tctctttaag tggctatgca tattttcact 300
 tccaccagca ataatggag ttctgttgc tccacatgct cactagcatt tgggtgtgtc 360
 agtgttctgg attttggta ttctaataag tacatagtca tatctcgttg ttttaattta 420
 caattcccta atgacatatg atgttgaaca tcttctcata tgcttatttg ccatctgtat 480
 atctactttg gtgaggatc tgttcagatc ttttgcctt tttttttctt tgagacagag 540
 tctcactctt gtcacccagg ctggagtga gtggcacgat ctcagctcac tgcaacctct 600
 gcctgctggg ttcaagcaat tcttctgcct cagcctccca agtagctggg attacaggca 660
 cccaccacca cgcaccagta atttttatat ttttcataga gatgggggtt cgccatattg 720
 gccaggctgg tctcaaaact ctgacctcag gtgatccacc tgccctagcc tccgaaagt 780
 ctgggattac aggcgtgaga caccacaccc ggctcttgtg cagctaattc tattttattt 840
 gagatggagt cttgtctgtg tgcccaggct ggagtgcagt ggcattgatct cggctcaccg 900
 caacctccgc ctctcaggct caagagattc ttgtgcctca gccttccagg tagctgggac 960
 tgtgcaccac catgctgggc taatgtttgt attttttagta gaggttgggt ttcacttagc 1020
 caggctgggt ccgaacttct ggccycaaaa gatctgccc cctcggcttc tcaaagtgc 1080
 ttggattccc aaagtgtctg gattacaggt gtgaaccatc atgactggca aagcatatgc 1140
 ttttgaggcc cattgtcttt cctaatttgt tgaatacata ctacatgagt atcttcaaac 1200
 actgagcaac tacgaaattt tttgtgaaat gccagtagaa atactaataa gtatttatatt 1260
 tccaggtaaa atgagacacg ggttttttaa agtactgaa tgtgcatgga agtatttttg 1320
 agactcacta aggaaataga ggcaccagca ctctctgtaa ttttttagta aagactccta 1380
 tctgagggaa tctgggatc ccccaaaaag gatctcagtt tgatcaccct acagtgaagg 1440
 tcaacaagtc ctaccaaga attcaaaaaca cctgtcagtc tttagtctcc tagtcttgaa 1500
 gtttgagcag agtcacatat taccagagaa ttcgaggata gtatctccga gaagccggga 1560
 aaaaactcag ttaagagaga agggatgctt taaaaaaaaa aaaagaggtc tttagacgtaa 1620
 atttctctta cctgaaagtc ataccatgat acatctcagt ggcctagcaa gtgtccagga 1680
 cagtggatca aagtggacct cctgccaaaa gcaaaaacat gttttgtata aaagatttaa 1740
 agaagaaaaa ggtaaaagctt cctgccaaaa caaaagacag agcagtgctc tcaaaattct 1800
 gaatcagaat ggcacagac tctccaactg caaaagacag agcagtgctc tcaaaattct 1860
 gagaaaaaat aatgtctaca ttaaaatttt tataatcatc ttcaattttt aagatactca 1920
 ctcaaaactat caaataaatg tgaggctaag taagagttaa gacctacaag gccttttttt 1980

15

gttttgTTTT taaaacgtct tttggatcaa tcatgagatg tagaatctaa taaaaccttt 2040
 ttatgattta tctcccataa accatttttt caggaaagta gataatatgc tccacaaaat 2100
 aaagtacatc aaggaaaaaa ggcatcttag tctttgatgg aagaagagg aagtccagt 2159

<210> 33

<211> 450

<212> DNA

<213> Homo sapiens

<400> 33

agaaaacaag atccagatac aaaaatcgat tgtattttta ctatgctaata aattagcaga 60
 tattgaaact ttttaaacat acaattttatt atagcatcag aaaaatggaa tgcttaagta 120
 taaatctgac aaaaaatgtg agctacctgt acactggacc actaaacact agtgaaacaa 180
 aattgaagag ctacttaatt ggaaatcagt ttccccccag atttatctat agagtcagtg 240
 aaatcccaat caaaatctca gcaagggtctt taagaaattg acaatcttat tttaaaattt 300
 aagtggagat gcgaaataac taaagcaatt ctctgacaaa aacaagaaaa aagctagaag 360
 gctaacaacc acactgattg caagatttat cagaacaggt ataataatca ggccagtgtc 420
 atatcggcat acacgataga ccaggagatc 450

<210> 34

<211> 584

<212> DNA

<213> Homo sapiens

<400> 34

ctagacttat ggatttgagg gagctgtgtg aaactcatca tggcaaatat gcttatgtgt 60
 atatatcctt tgccatacat gtgctgcaaa ctgtaatgaa atgttattta taagactggg 120
 aaggcatgtg ttattagact ggacacacaa aagcccttga ttatctagga agcaatcctc 180
 taggggccag atgtagtttg gaatgtgggt gtttagtata actgtacttc attactgatt 240
 tttatttcta tgctgtttga ctgtattagc tctttgttat tattggggag gtagccagag 300
 gtctccagat tcccataatg aatttacagg tgtgatctta tggacaagga ggagtcagct 360
 gtattagtgt ggggttcaat ctgacctgat aagcttttcc tagttgggtt tacagatacg 420
 agccctgata tactccctgc tgccactgtc tgtttctatg atgcatgtca ccatgatata 480
 tgagtatgta tgaaaatata tttaggctaa ttttaactag aatatggaaa ggaaaaagtt 540
 ctattgctct gcattgctct gttttcagca atcactgttt ttca 584

<210> 35

<211> 642

<212> DNA

<213> Homo sapiens

<400> 35

gctagactta tggatttgagg ggagctgtgt gaaactcatc atggcaaata tgcttatgtg 60
 tatatatcct ttgccataga tgtgctgcaa actgtaatga aatgttattt ataagactgg 120
 taaggcatgt gttattagac tggacacaca aaagcccttg attatctagg aagcaatcct 180
 ctagggtcca gatgtagttt ggaatgtggg tgttttagtat cactgtactt cattactgat 240
 ttttatttct atgctgtttg actgtattag ctctttgtta ttattgggga ggtagccaga 300
 ggtctccaga ttcccataat gaatttacag gtgtgatctt atggacaagg aggagtcagc 360
 tgtattagtt gggggttcaa tcttgacctga taagcttttc ctagtgggtt ttacagatac 420
 gagccctgat ctactccctg ctgacctgt ctgtttctat gatgcatgtc accatgatata 480
 ctgagtatgt atgaaaatat atttaggcta attttaacta gaatatggaa aggaaaaagt 540
 tctattgctc tgcatttgct ctgtttttca gcaatcactg tttttcacc acatatagaa 600
 agtttgaaag ctctctctga tgtctggcaa ccagatctcc ca 642

<210> 36

<211> 669
 <212> DNA
 <213> Homo sapiens

<400> 36
 ccaaaattta ctagaatgtc ctgaaccaca tctttcataa tgttgctgac tcaaagactc 60
 ttgaaggctc ctgaccacat tattcgcaat tctaactctc ttgccacccc ttccccatga 120
 cccatgtaca attacatgct ctagatcttc tcctcaaaga tgaacataag tctgaaatat 180
 caacaccttg gcagccctat tatcaattgc tgatctgtag tccccatgta agtacgcctt 240
 ttttagcaac cagtttttgc cccagccata ttaatacttg tggtcagtggt ttaacaatgt 300
 tgaagcttaa attatatcca gatgaaatct taaaaggaa tcacttgtgt tcctctgtgt 360
 taacacagga atcccagcat gtgtttctct tccaggaaac cataattata tgtacaaata 420
 tctacccgga caattagggg cataatcatg ctctaaatag aagtgttcaa acaagtcaac 480
 accttctctc cagttattcc tctttctctc ttctcttaga tgtcatgggt tctgtgtctc 540
 aagacattta tgatttgatt tttctaacc tttctaggtt ctattagagt caattagaca 600
 acatattcct tctttctaag aatctggaca aggaggtata cttttctaaa ttttaacctt 660
 attaatgcc 669

<210> 37
 <211> 1006
 <212> DNA
 <213> Homo sapiens

<400> 37
 tcttaaaatg agcaccctca ggactgtag gtaggagagg tgtagattt caagtagata 60
 caaataggct cagaaggtaa aatgaggacc caaggataga agagcgacag tgatttcagc 120
 tgagcctcag ttccaagcac agaacttttc agaaacagaa tgggttgcat aatatgtccc 180
 cttttaaaag acactttgca gacctggatg cctgtgtgtt ggcatggagc atagaggttt 240
 cctgtcctgg gtaaaccatgc tgtgctggac taggttctct ctgaaagtct ctccctgctt 300
 caggagtcta gaattctaag tttcttctca ggagactcca aaatttacta gaatgtcctg 360
 aaccacatct ttcataatgt tgctgactca aagactcttg aaggctcctg accacattat 420
 tcgcaattct aactctcttg ccacccttc cccatgaccc atgtacaatt acatgtctta 480
 gatcttctcc tcaaagatga acataagtct gaaatatcaa caccttgga gccctattat 540
 caattgctga tctgtagtcc ccatgtaagt acgcttttt tagcaaccag ttttgtccc 600
 agccatatta atacttgtgg tcagtgggta acaatgttga agcttaaatt atatccagat 660
 gaaattttta aaaggaatca cttgtgttcc tctgtgttaa cacaggaatc ccagcatgtg 720
 tttctcttcc aggaaccat aattatatgt acaaatatct acccggaaca ttaggggcat 780
 aatcatgctc taaatagaag tgttcaaaca agtcaacacc ttctctccag ttattcctct 840
 ttcttctttc tcttagatgt catgggttct gtgtctcaag acatttatga tttgattttt 900
 ctaacccttt ctagggttcta ttagagtcaa ttagacaaca tattccttct ttctaagaat 960
 ctggacaagg aggtatactt ttctaaattt taatcctatt aatgcc 1006

<210> 38
 <211> 589
 <212> DNA
 <213> Homo sapiens

<400> 38
 aggagctggg ttttgcttaa cagaaggagc actgacccat gttatagaca atcgcagaat 60
 ttcatatccc catctataaa atgaaaacac aatacttctc accaactt atacagcacc 120
 tactatgtgc taggttagag atcataaact ggtgatagt aagtggata taacctcag 180
 acttggtctg tgtgttctac gcagttgatc tgcaccagcc tttgttaaaa ttggaaggaa 240
 attgctaata tttaaaatca ggatatttcc cagcaaaatc tacatttcta gtatctcaga 300
 aaaatcatta tttggcagca ctgggcccaga atttctgcag ggcaattgtt gtctgactt 360
 ggggtggctg tggaaatggg cgtgtactcc taagtgtgtc ccaattgcta ccgctctatt 420
 acttcatcct ttaatgttca ctactcttgg ccctgtggga tttttgaggc tgagattcct 480
 atattaggtt ctgaaggcaa aacacacaca gaaaagaat atttcaggcc cttcctgagc 540

atactcatga tgtataactt ttatgacagt aatagtagta tctagcaat

589

<210> 39
 <211> 528
 <212> DNA
 <213> Homo sapiens

<400> 39
 aagacctgtc tttatTTTTa gaagtaagaa taaaagagat tgtggtggag tatcacaggc 60
 agcgtgggag cactgaggga gccctgacc caccctagga gtggatcagg atgacttctg 120
 aaaggccaaa ctgattaata agggataaat aaagtcatgc aaatgaaaag gttgtatatg 180
 tgttggggga aagcattcca gacagaagga ccagtgtgtg caaaggccct ggggtgagag 240
 gtgcctaata agtactgaat atacaaagag gtagagctgg gactaaacca ctgtgtctac 300
 tttgctgtct tgaattccga ttccaaggag tggaaatagac ttcaaagtgc ttcaagtcca 360
 cttgtttctg ccaagttctc atttttgttc catgaaggca gagcaccttc tttatttcat 420
 ccactgatga cttctcagcc tctagaattc tgccttatga tggatttctc agaaatatgt 480
 ttgtgtaatg aagacaagga cagtggtagg agtttacatt ctactggg 528

<210> 40
 <211> 673
 <212> DNA
 <213> Homo sapiens

<400> 40
 caaaaaataa aaacaaaac attagttggg cgtggtagtg tgtcccaggc actcaggaag 60
 ctgaggtggg aggattgctt gaggccccga gttggatgct gcagttagct atgattgtgc 120
 cactgcagcc tgggtgacag aacaagaccc tgtctttaaa aacaagaagt aagaataaaa 180
 gagattgtgg tggagtatca caggcagcgt gggagcactg agggagcccc tgaccaccc 240
 taggagtggg tcaggatgac ttctgaaagg ccaaactgat taataagggg taaataaagt 300
 catgcaaata aaaagggtgt atatgtgttg ggggaaagca ttccagacag aaggaccagt 360
 gtgtgcaaag gccctggggg gagaggtgcc taatcagtag tgaatataca aagaggtaga 420
 gctgggacta aaccactgtg ctccactttg ctgcttgaat tccgattcca aggagtggaa 480
 tagacttcaa atgtcttcaa gtccacttgt ttctgccaag ttctcatttt tgttccatga 540
 aggcagagca ccttctttat ttcatccact gatgacttct cagcctctag aattctgcct 600
 tatgatggat ttctcagaaa tatgtttgtg taatgaagac aaggacagtg gttagagttt 660
 acattctact ggg 673

<210> 41
 <211> 447
 <212> DNA
 <213> Homo sapiens

<400> 41
 ctcaagcagg gctagcacct ccaatctaga gcaccctgca cttccggctc caccgggtctt 60
 cttgtccctt cactgccttg cctaggggtg cttctctctc ctctcttaag ctgagtacaa 120
 gtgataatat agtgattaac acaatgctgt agtgttttcc tggtaaagag ggaatgggtg 180
 attttccagg agaataaaaa atgaaattgt cattggagga cctcctcagt tgaatacatt 240
 ctgtggctga ttctctccta ttttgttttt tgggtggttg ttggtttttg ctttttcagt 300
 agctaccag gtatacaaat agcttctttg cagttctgat catctttagg ggccgcattg 360
 ggcataattg gaataataat actagctaac ctgcttgcaag ggcttgctct gtgctgtgca 420
 ctttgtgagc actttaataa taggagc 447

<210> 42
 <211> 562
 <212> DNA

<213> Homo sapiens

<400> 42

```

ctcaagcagg gctagcacct ccaatctaga gcacctgca cttccggctc caccggctctt 60
cttgccctt cactgccttg cctaggggtg ccttctctc ctctcttaag ctgagtacaa 120
gtgataatat agtgattaac acaatgctgt agtgttttcc tgttaaacag ggaatgggtg 180
atthtccagg agaatagaat atgaaattgt cattggagga cctcctcagt tgaatcatt 240
ctgtggctga tttctccta ttttgtttt tgttggttg tgggttttg ctttttcagt 300
agctaccag gtatacaaat agcttctttg cagtctgat catctttagg ggccgcatg 360
ggcataattg gaataataat actagctaac ctgcttgca ggttgctct gtgctgtgca 420
ctttgtgagc actttaaata taggagccaa acctctctt ccaaaagcct gaagggcagg 480
tgtctcgca gttcccatc catagatcac catccttcca tggaaagtac tctgtggact 540
gtaacttgcc atctagactt tt

```

562

<210> 43

<211> 848

<212> DNA

<213> Homo sapiens

<400> 43

```

gggtctttct agctttcttg tctttgtga agctggactg gtgatgtgca gttgaagaca 60
gcatcatcgg gggccttctg ctccatgtgt accctccagt atthgcaaaa gattgaacct 120
acaagatacg ttattagggc aagtatttac atggaaaggc tctgagttct ccaagacttt 180
ggtcattttt tacaagatga tgtactaccc tgatgatttg tggaaatctt ttaggaaccg 240
tgactgtggt gcttttctga tcatgggtac agggccatct ttgttgaggc ttcccatgtg 300
tgtgggcaca gagcttctgt ggcattccag cagtagatta atggagctgt catcctctga 360
agcctcatgg gttgtgcatg caaacctggt cctgtgaact gcatgggagt ctcttaaaag 420
ggcagagggg ttcccttctt tgtgaaagg ttagaatggc acatatattg aatttccaga 480
ctcatctttt cccactctca cattactct gtatttgccc gtactaaatt gttgacagtt 540
ctccaaatac aacagcattg ctattctgct gccttcgtac atgccgttta cttactgtc 600
acattgtcca ggaattcatc cctgccatga ctgcagtgcc ccctctggga gctccccgtg 660
ccctgtgcct gccgtgtc gagcttccag catgctgggc tgtggagggt ttggtctgtt 720
tgccaccca gcaagcctct aagctcctca aagacaccaa ctgtcacgca tatctggagc 780
agcacctggt accttacggg tctttaaag ccggtgaat gaatgatgtc tctgtctct 840
ttaaacc

```

848

<210> 44

<211> 1111

<212> DNA

<213> Homo sapiens

<400> 44

```

gggtctttct agctttcttg tctttgtga agctggactg gtgatgtgca gttgaagaca 60
gcatcatcgg gggccttctg ctccatgtgt accctccagt atthgcaaaa gattgaacct 120
acaagatacg ttattagggc aagtatttac atggaaaggc tctgagttct ccaagacttt 180
ggtcattttt tacaagatga tgtactaccc tgatgatttg tggaaatctt ttaggaaccg 240
tgactgtggt gcttttctga tcatgggtac agggccatct ttgttgaggc ttcccatgtg 300
tgtgggcaca gagcttctgt ggcattccag cagtagatta atggagctgt catcctctga 360
agcctcatgg gttgtgcatg caaacctggt cctgtgaact gcatgggagt ctcttaaaag 420
ggcagagggg ttcccttctt tgtgaaagg ttagaatggc acatatattg aatttccaga 480
ctcatctttt cccactctca cattactct gtatttgccc gtactaaatt gttgacagtt 540
ctccaaatac aacagcattg ctattctgct gccttcgtac atgccgttta cttactgtc 600
acattgtcca ggaattcatc cctgccatga ctgcagtgcc ccctctggga gctccccgtg 660
ccctgtgcct gccgtgtc gagcttccag catgctgggc tgtggagggt ttggtctgtt 720
tgccaccca gcaagcctct aagctcctca aagacaccaa ctgtcacgca tatctggagc 780
agcacctggt accttacggg tctttaaag ccggtgaat gaatgatgtc tctgtctct 840
ttaaaccac cttctactat gctaccataa tggatatttc ttctaactgg caattttaaa 900

```

19

```

gatcctgctg tggccttttg tcaggctttt gagcagggtt tggcaaatcc gtggcctatg 960
gaccagggtc ggcccgcggc ctgatgtgca tccttgcgct ggccggttca ggatgaattt 1020
acagttactg acaccaattc ctgtggaaaa taataaaaga ctgcggttct tcacatcacg 1080
tagcttaaaa agggaacacg gggacaaact g                                     1111

```

<210> 45
 <211> 626
 <212> DNA
 <213> Homo sapiens

```

<400> 45
tgttctgaca tcaacaggaa aaatggtaca agaatatttt cagatcatgc caaaaagcag 60
cacttcgtta aaaggaagaa aaaatttcaa gtaaacata aacagggttt tagattgctc 120
gataattcaa ttagtgaatc aaacaatgat aaaagctata tatttcctgc tgatttgta 180
ggaaatagtg acactgacaa agatagcatt acctaagaat ataaaagcaa agatagcgtt 240
gccacagact gcttaatgtg tgtcatctat caaaggggta tatgtgatga gaagaaaaac 300
ttgaaatgcc ctcaaagtgt tcagctatca gaaactgaaa aaactcttac tagtgtgttc 360
cgcataattg tgagcaatat tctaaagatc gacgtttctt cagttatgat tttcttgagg 420
ctacatcaga gaacttcctt aaacctgtcg gtaatacaaa atcagtgcgt catggcaaaag 480
gggagacatt atctatctgt tcttgactat ggaaaataat gttgcagaat ctttgtcctg 540
tgtgtgaaga agcgatgagt acaggaccag aactgtccgg aagacgtatt tcaggagacg 600
cacatggcag tcgggcgcgc ctctag                                     626

```

<210> 46
 <211> 185
 <212> DNA
 <213> Homo sapiens

```

<400> 46
gaagaaactg tgaggtcaca atacttttga ttcattatgt gaatatacat acacactcac 60
atctctatta ctgtatccat ctctatatac ttgaactcca tatgctctat attaacttcg 120
ccaaatccaa cccaacaaac agggttcctc tctgattttt ccccccatat ttatgattct 180
cagac                                     185

```

<210> 47
 <211> 268
 <212> DNA
 <213> Homo sapiens

```

<400> 47
atggatttgc cacaagctgg ctttgaaagc agtggtagag tgtgaaagaa gttaccttaa 60
gacttcttgc cagttgcact gtaggtacga tgtactgttt gttgtgattt gactttcctc 120
caccaccccc ctgccccagg aagatgtgat cttgtgcac ttgtgttcac gcagagtagg 180
gtagttggat ctttgtcaag tctcagtgat ccacatgcgt gcatctattt tgtcagtcgt 240
cttgtctttg tatccatgtc ataactgtc                                     268

```

<210> 48
 <211> 108
 <212> DNA
 <213> Homo sapiens

```

<400> 48
gtcgacgacg acagcaatgc cgatccgcgt caccgccgca accggctgcg gctgcagggtg 60
atgcctgccc tgcgcgaggg cttcccgcag gcgccgctgg cgctggcc                                     108

```

<210> 49
 <211> 83
 <212> DNA
 <213> Homo sapiens

<400> 49
 gatcgagatc ggcggcgtgc cgctgggtgca tctgcccgcc gaggcggtgc gcgcgccctg 60
 gccgctcgac gagcgcgagg tgc 83

<210> 50
 <211> 475
 <212> DNA
 <213> Homo sapiens

<400> 50
 aaagaaacaa gcaacaaata ggaaaatcaa attttttagaa gtaggtgcat aataggggaa 60
 tagcttaagg ggagaactat gatgttaatt ctttgaaagt gagtaatgta attagaacaa 120
 taacactatg agtttttcta taaacaaaat atagcaagat taagttgata acatacattt 180
 ctaaaatttt ggcttcctta gagaaagcca accaaatata aaattttaca gcagagtcaa 240
 gttttttcag tttggcctat attttctttg gtaacactgt tctgaatgta tatgcagtgt 300
 ttatttcaca acttcctctt gaatgacctt tcaaaaatta atgattcttc acattcatga 360
 ccagatgttt tctctgatgg aagcatctga tgtttgagc catcaaataa gattcaaaat 420
 gtctgtttca agcaaatcaa gtaaaacttc tccatcacat caaaagtaag gcttg 475

<210> 51
 <211> 607
 <212> DNA
 <213> Homo sapiens

<400> 51
 aaagaaacaa gcaacaaata ggaaaatcaa attttttagaa gtaggtgcat aataggggaa 60
 tagcttaagg ggagaactat gatgttaatt ctttgaaagt gagtaatgta attagaacaa 120
 taacactatg agtttttcta taaacaaaat atagcaagat taagttgata acatacattt 180
 ctaaaatttt ggcttcctta gagaaagcca accaaatata aaattttaca gcagagtcaa 240
 gttttttcag tttggcctat attttctttg gtaacactgt tctgaatgta tatgcagtgt 300
 ttatttcaca acttcctctt gaatgacctt tcaaaaatta atgattcttc acattcatga 360
 ccagatgttt tctctgatgg aagcatctga tgtttgagc catcaaataa gattcaaaat 420
 gtctgtttca agcaaatcaa gtaaaacttc tccatcacat caaaagtaag gcttttatatg 480
 gttcacaagt agctatatga aataaacaga atttaaacga tcttaataat ttttttcttt 540
 aaacaagggtg acaaaataac aatgccaata tataaaaact cctcattaat gataagtgtc 600
 agatgga 607

<210> 52
 <211> 590
 <212> DNA
 <213> Homo sapiens

<400> 52
 ctctcatta atgataattg ctgatggac accatgtaaa gtatggaaaa tgctgtctg 60
 aacaaatgct tttgctaaat tctctgaatt ttttttgtt tttctcacc agttagcttt 120
 gatgttttga tcagagtttt tagaaaaatt ctaggatctg ttgccttttg acttttagagc 180
 ttcttgagac cacatgtcag tactaaaacg ttttcttaag ccctcgcttt ccatagcaaa 240
 aacatgttat gtccattatc cacctaactc atacttaaaa acaacacca agatgctcta 300
 ttttgttttc aaagtcagag aagaaaaatag aggggaagta tttttatgtt cttttccctg 360
 aattggctga agctagttag ttcaaaaaag atacaaaata tggaatacca cctattttat 420

21

```

ttcctggcaa ctgtttcatt caaatcatag agtaacatat gattttactac actcctttat 480
gaatattaat ctcgatatctt cacagaatga cttaatatca ttgatcagct agaacatcga 540
cctcacctgt ctgttggtttt taacgaaatg tttattccta gtcaaaccac 590

```

<210> 53
 <211> 217
 <212> DNA
 <213> Homo sapiens

```

<400> 53
agtctgctaa ctcatccag tggttttttc caactgcac tcagttatct tacatagact 60
gcaagaagtg agaaagacaa gaggttatct agtcagcct tgctatttta tagtttaaat 120
ccctcaacca catccctgat gaacttttgc cagtgccggt aattaacaat atcacaaggc 180
tgttctgatt gtctgtattt ctcagtgttt gttagag 217

```

<210> 54
 <211> 430
 <212> DNA
 <213> Homo sapiens

```

<400> 54
aataaagata agaatgacaa cagattttctt tttgggaaca atgagagtgg gaagacaatg 60
agcaacatct ttaaagtact gaaaggtatc agcagaccca tgctacaaaa aatgtaaaag 120
aacatcatca ggcagaagga aaaaaatagt atcagattga agtctgttct acacaaagta 180
atgaatacca gaaatgataa ctacctgggt aaatatataa gattattttc ttcttattta 240
aagtaagagt gagattctta tcaacaatag cataaaggct gaaggggaga aatggaagtc 300
tattagtgtg atcttatata tgatgtggta tgatgtcact tgaatgtaga attataaaga 360
taaacagcat aaactcttaa agcaaccacc aaaataacaa agagttataa ctaataattc 420
agcaaaggag 430

```

<210> 55
 <211> 2956
 <212> DNA
 <213> Homo sapiens

```

<400> 55
gttggtgttg ttttttttga gacagagtct tgctctgtcg tctaggctag agtgcagtgg 60
cgccacctcg gctcactgca acctccacct cctgggttca agtgattttc ctgcctcagc 120
ctcccgagta gctgggttta cagggtgctcg ccaccacgcc cggctaattt ttgtttcttt 180
agtagggttt caccgtgttg gccaggctgg tctcgaactg ctgacctcgt gatctgccc 240
ccttggcctc ccaaagtggg gagattacag gcgtgagcca ctgcacctgg ctttttattt 300
ttttaacttt gtatacggta ttttcttttt ctgtatagaa gtcaaactat tttccttcat 360
ggattctggg ttttgtctct tcattccaag accatttaaa aaaatgtgtt cacattttcc 420
tctgatactt ttaagggtgtc tttctgaaga taaaacctga tgtgtctgca atgctagagt 480
gaggcttgag tatgggcaag ctctctgagt gcacgtgtga gctgaggaca gcatggcgtg 540
tgaggaaagg tcagtccaca cagctcatgt aagctcacga gagaggctac tggcttcact 600
gcacgtgtct actgggtgtt ttgacaacgt ggagtgaata ctcatgtcc tcacaaattc 660
aaatgtctgt tttatcatgt ataaatatta tattggaaaa aaataaaatc ataataag 720
tatttgctca cttatottga agaaaaacac atacatgttg cacttctgaa tttaccttaa 780
cctgtttaat acctactgag aaagtctact attcagaatg cagaaaaagg tggaaggagt 840
ggttagggcc ctaaaagtca aactgggtcc ccgcagccca gagatcaaca ttatttaaaa 900
actcaccatg caaagctaag agagaacgaa ccatgtaacc ctttttgaac tattacattt 960
tcaactcaaa gcttggccct atcttccagt tacacgtcta taaatgtcaa ctacgaagcc 1020
tttcagaggc cttacacttt gcaaatgaag tcagtgaac cctcctgcac acagacagag 1080
cccaaaggac aggagtgcag ctggcagtgc agcccttggt ggggccaagg ggcaggtcac 1140
atggaagggt gcgggttcct cccatgtcca tacgtgacc cctcactcat gctccagac 1200

```

```

ccctctggac accgtgctgc tggcagatgc tgtgctcctg ggaggtggga tgcaagctga 1260
accttgctca ctcccttttg gctaaatgac aggtgagcac tgggcacagc aaatgtgact 1320
ggccacagcc tcctctgcag gggcaacaag ttcccacac aagatcccggt taccatccca 1380
cacaccccggt ctccatctct ctggatcctt gttcagacac agtgttttta tcaacaccca 1440
cagagggaaaa tgggtaaatg cgaaaactcg tttttgcagc tttaaattac ctatgtcctc 1500
agaatgtagc agaattcaca gctggctggg aaaagctata atacatgcac tgcacacact 1560
aacgcgtttg aatataaata agcgtatctt taagttctgt aaagtccctt accgccaaagt 1620
agaataaaga caccaacctc ttttgcctat aggtcctcctc tctcctctgg ataccgttct 1680
ataatctgaa gtaatctagg aaacttcaat ctggcttcat tggaaattaa ttttaaagct 1740
ttcaacattt tctccaccac aagtgcctgga tacgcctgca gttctgcaga atcaataact 1800
atcaaggaca ccaagaaga aagcaatggg caatgtatcc caatatccat aaactatgat 1860
gttaaatgct aacactttcc ctttttggct tgtattttgt agtgtcattg ttctcttctt 1920
aactaccact ttacaccaac aaacaccagg tacagttttg tatctatcct ggagccaaat 1980
ccttccatta gagtgcctat tctgcatgaa gcacagtttg aatcctgggc tgggaacata 2040
aggggcaatt ggtggttatt gaatttattc caggagcatg aagcaggcca cagcagccag 2100
taatatgaa gctgcaagca aaatatcaaa gtagaaatta aacaaatgga aacagaggac 2160
cacttgactc catttaaatg taggtcatgt tgccttagaga ggccattgtc tctctctttt 2220
ttttttttta agatggagtc tgcctctgtc acccaggctg gtgtgcagta gtggatatcg 2280
gctcactgca acctctgcct cctgggttca agcaattctc ctgcccagc ctctctagta 2340
tctgggacta caggcatggg ccaccacgcc cagctaattt ttttgattc ttagtagaga 2400
ggtgggttca ccacgttggc caggctgggc ccgaactcct gacctcaagt gatccacctg 2460
ccttagcttc ccaaagtgtc gggattacag gcgtgagcca cctcacctgg cctaatttca 2520
ttttatctcc tttgtgaat tattagttat aactctttgt tattttgggt gtgtctttaa 2580
gagtttatgc tgtttatctt tataattcta cattcaagtg acatcatacc acatcatgta 2640
taagattaca ctaatagact tccatttctc cccttcagcc tttatgctat tgttgataag 2700
aatctcactc ttactttaaa taagaagaaa ataattctat atatttacc aggtagttat 2760
catttctggg attcattact ttgtgtagaa cagacttcaa tctgatacta ttttttctc 2820
tctgctgat gatgttctt tacatttttt gtgcatggg tctgctgata cctttcagta 2880
ctttaaagat gttgtcatt gtcttccac tctcattgtt cccaaaaaga aatctgttgt 2940
cattcttacc tttatt

```

<210> 56
 <211> 517
 <212> DNA
 <213> Homo sapiens

```

<400> 56
cctggctgga gcggacacgg tcaagaccgt cctccctacc ttctcccttc aacccaagct 60
caactcaacc aaaaatggcc cctctgtccc catgcctgat aggaaagtca ggggaaagtc 120
tgtccgatta ctgtcaaaga agacaggagg taagggtcag agtggaccac tgactgaata 180
tgagtcgcag aagtgttaga ggcagaagtc cagggccatt tccttaatat cgaagtgtct 240
ctgctggagg tctgggatgg atttttgccc tgcatttaga agttctgggg tcctgggaga 300
ggggagagaa gcccaatagc agaggagaca gagtgtgggc ggggcgagcc ggaggggtgc 360
atcctgggag agcaccaggg tgagggaggg gtgaagatga gcccgcagc ggaaagcgtg 420
gcgagtgtgg gaagtcacct gcccctcggc ctgtgagctg ctctgcttgg agtgactaag 480
gctcgggagg tccaggctcg gccagaggca gctcata

```

<210> 57
 <211> 1490
 <212> DNA
 <213> Homo sapiens

```

<400> 57
ggggaaccag acgcccagtc acaggcgaga gccctgggat gcaccggcca gaggccatgc 60
tgctgctgct cagccttgcc ctctctgggg gccccacctg ggcagggagt aagtcagtgg 120
ggctgtccct caatctcccc tgctccctc caggagagcc agggactcac ccggcccttg 180
tcccagacta actctgtgta cagaaccatc ctgtctgctt ggagggggcg ggtccctctg 240

```

23

```

tctggcagag gtcaccccca taccacgca tggggatttt cttccctttg ggtctctctt 300
ttcttcagag atgtatggcc ctggaggagg caagtatttc agcaccactg aagactacga 360
ccatgaaatc acagggtctgc ggggtgtctgt aggtctttctc ctggtgaaaa ggtgagtagg 420
gctatggtca tgggcccagc gccatgtccc ctcccatccc acagtttcag gaactcaggg 480
cagcgggtaa gcaccgtgg ccacttttgc cacacatgcc tggctactgt cgatgcttcc 540
tggctcccgc tgatgcttcc tggctggagc ggacacggtc agaccgtcct ccctaccttc 600
tcccttcaac ccaagctcaa ctcaaccaa aatggccctt ctgtcccat gcctgatagg 660
aaagtcaggg gaaagtctgt ccgattactg tcaaagaaga caggaggtaa gggtcagagt 720
ggaccactga ctgaatatga gtcgcagaag tgttagaggc agaagtcag ggcatttcc 780
ttaatatcga agtgtctctg ctggagggtc gggatggatt tttgccctgc atttagaagt 840
tctgggggtcc tgggagagg gagagaagcc caatagcaga ggagacagag tgtgggcggg 900
gcgagccgga ggggtgcatc ctgggagagc accagggtga gggagggtg aagatgagcc 960
ccgtcagggg agcgtggcg agtgtgggaa gtcacctgcc cctcggcctg tgagctgctc 1020
tgcttggagt gactaaggct cgggagggtc aggtcggcc agaggcagct catatgtggg 1080
ccacagtgc ggcagctggt gccttctggg tcacggagac ctggcgtgc acgcagctct 1140
cctcaccagg atctcagtga ctctcccaa aagtcacacc cactttgcag acggggaaac 1200
tgagtccgga gaggtgggt aacgagctca agatcacagg gccaaaagt ggtagaatca 1260
gggttggtga ccagtgaatc tgtgtcagag acccaaagtc tgatggtgct ggactctctg 1320
catcccgga agggagatgg gggcgctgag gacccgggat gtgctgggcc atcccagatc 1380
tggacgtcca aagctttgcc tctctccag tgtccaggtg aaacttggag actcctggga 1440
cgtgaaactg ggagccttag gtgggaatac ccaggaagtc accctgcagc 1490

```

<210> 58

<211> 436

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (197)

<223> a, c, g or t

<220>

<221> unsure

<222> (432)

<223> a, c, g or t

<400> 58

```

ctctgtctcc tcataggaat ttcttagttt cttggctttc gaatgtgact caaccctcc 60
cttggcctgt ctgtctgctg tgtcgctttt aggttctgct gccacggcta actatgtttc 120
cctgtgtttc cagataaact tgtgagggtc agaagctgac agaccaagct catttttcaa 180
gccaatctgt gtcatanaga gaccacgggt tttccttggg ttgggtcctt ctacctggtt 240
cagtcagctg tgaacaaaac ttgtggaatt tggatcattt ccttaaaatg gagatacag 300
agatcaccat ggctggcgtg aaactagttc tggatctgat tgtcttttca attgtttgtc 360
catcagggtg acccactctg aagggaactt tggtaacatt ttccccaaaa taaagatcat 420
taattaatta tnaaaa

```

436

<210> 59

<211> 458

<212> DNA

<213> Homo sapiens

<400> 59

```

ctctgtctcc tcataggaat ttcttagttt cttggctttc gaatgtgact caaccctcc 60
cttggcctgt ctgtctgctg tgtcgctttt aggttctgct gccacggcta actatgtttc 120
cctgtgtttc cagataaact tgtgagggtc agaagctgac agaccaagct catttttcaa 180
gccaatctgt gtcatacaga gaccacgggt tttccttggg ttgggtcctt ctacctggtt 240

```

24

```

cagtcagctg tgaacaaaac ttgtggaatt tggtcatttt ccttaaaatg gagatacgag 300
agatcaccat ggctggcgtg aaactagttc tggatctgat tgtcttttca attgtttgtc 360
catcagggtga acccactctg aagggacttt tggtaacatt ttcccaaaa taaagatcat 420
taattaatta taaaaaaaaa aaaaaaaaaat gagcggcc 458

```

```

<210> 60
<211> 359
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (336)
<223> a, c, g or t

```

```

<400> 60
cggacgcgtg ggaacacaaa actgcatcat ccaaaaatac acctttgggc cacggatgcc 60
actggaagac atctgaatth tagacctcca gagagaagat ctgggtgggt agctccagag 120
tggaggcatg cttgcttttt ctttacactt gtgaagagga atggatccgg acatctgcaa 180
tctgggtaga ggacggcagg cagcaagctt agccaactcg ccaggcttct cagcccttac 240
tctagacatg tgatccttcc tccacgtgat atacttcaca actttcttac ggctactcaa 300
ggcatcccaa gttaaaagga aggtcagatg tgattnatca ctttattatg ataaaaaaa 359

```

```

<210> 61
<211> 932
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (161)..(180)
<223> a, c, g or t

```

```

<400> 61
tggccagaga catatgaaaa gatgccttag acatatagca tcttttctca tccacttact 60
aggagaaatg ctcactaaaa ttatcctgta atgccattta aaaaaatctc agattgttga 120
agtacaaaaa gttagataac atattatcaa ccaaaatgtg nnnnnnnnnn nnnnnnnnnn 180
ttgggccagc tgtgtttggg taaactagtt aagggtgtag ggtgtttgg tcaggaatta 240
aatcataaag aaaaacaaaa cctctgaaat gaaaactcat ggtgagggtg aaacttcacc 300
ccttgtagtc acttatgttt aactggtcta ctggattttt ttaaagggtt agaaaacaca 360
aactgcatca tccaaaaata cacctttggg ccacggatgc cactggaaga catctgaatt 420
ttagacctcc agagagaaga tctgggtggc tagctccaga gtggaggcat gcttgctttt 480
tctttacact tgtgaagagg aatggatccg gacatctgca atctgggtag aggacggcag 540
gcagcaagct tagccactcg gccaggttc tcagccctta ctctagacat gtgatccttc 600
ctccacgtga tatacttcac aactttctta cggctactca aggcattcca agttaaagg 660
aaggtcagat gtgattctca ctttattatg ataaaaaaa ttactattta aatactataa 720
ataaatatta taataaatac taagctagaa ccatacagaat acatcacttc tgtatccagt 780
tttcaaagta tctttggtgt ttgtcaggaa taaataaaag taatcatttt atttctatta 840
aattatatct ggcactagtg gctagtactt ttgtacttat tagtacaacc taaaaaagtc 900
ttaaaaagat ttcttttggg ttcagaacat aa 932

```

```

<210> 62
<211> 554
<212> DNA
<213> Homo sapiens

```

25

<400> 62

```

ctggcagatc cggacgggca ggactgggtg tgtcccatga gagcacctcc ttcttggcct 60
ttcttgtgga ctttgtccca caccacctgc ctgggttccct tccttttagtc acttccagct 120
ccaggcacag cagttggtga ctcccttgggt ggagccgtgt cccaccgggt cctgatactg 180
ccgtcttctc tttcacagtc ctccaggctt gggccagcct tgggggcagc agagcttctg 240
gggtgagtggt cgagatcctg tgtcctgaga gcggtagtca gggagagggc tggtcggggc 300
agggctgccc gggcaggaca caggatgcgg ccggccaggc tggggccaag gtgttcagac 360
ctggactttt ggctcgtgct ttcttcatgg ttgcgccttg ctgcgtgtcc cttggagtct 420
tcattttggtt ttgctttttt tgtttgtttg ttttcacctt atttttgcca gacttaagct 480
agtttttgctg ccttttgaaa ctagtggaa aatcatttta tttcctgggg ataatttggg 540
ggcttttgaa tcca 554

```

<210> 63

<211> 786

<212> DNA

<213> Homo sapiens

<400> 63

```

ccagtggcct gtgtcctagc aaatgagagc caccctgaaa aataaaatcc tgtctcccca 60
acgccagccc tggcaaggca cccagaactc tccggaatgc ttgaaggcag ggcctggcct 120
ttccatgggg tccagggtctg tggggtccct ggcggtactg tgggcctgca gagcggggca 180
tgtgggtgga agaccgtctc cccaccatgg tgggaaggga caaagggtgg ccttggcaga 240
tcgggacggg caggactggg tgtgtcccat gagagcacct ccttcttggc ctttcttctg 300
gactttgtcc cacaccacct gctgggttc cttctttag tcaactccag ctccaggcac 360
agcagttggt gactccttgg tgggagccgt gtccaccctg gtcctgatac tgccgtcttc 420
tctttcacag tcctccaggc ttgggccagc cttgggggca gcagagcttc tgggggtgagt 480
gtcgagatcc tgtgtcctga gagcggtagt caggagagg gctggtcggg gcagggctgc 540
ccgggcagga cacaggatgc ggccggccag gctggggcca aggtgttcag acctggactt 600
tgggctcgtg ctttcttcat ggttgcgcct tgctcgtgt cccttggagt cttcatttgg 660
ttttgctttt tttgtttgtt tgttttcacc taatttttgc cagacttaag ctagttttgc 720
tgccttttga aactagtgga agaatacttt tatttcctgg ggataatttg ggggcttttg 780
aatcca 786

```

<210> 64

<211> 575

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (411)

<223> a, c, g or t

<400> 64

```

ggcacagcta gttggtgact ccttgggtgg agccgtgtcc cacccggtcc tgatactgcc 60
gtcttctctt tcacagtctc ccaggcttgg gccagccttg ggggcagcag agctttcttg 120
gctgacatgg ggctcattgc tctttctctc aagccctctg agggacatca aaagcgtggg 180
acgcatccac tttccacca tcttggcttg cccactgtt cctccatcc tggagggcct 240
tccttaagca catgtgtggg ggtgggcagg cacactggct gatagctgtg gatgcggccg 300
tgacatcctt caccctgcc cccatggcat gcatgatcca ttagggagga ccgtctgcac 360
aaaggctctt tgccctgtgc aagcttcctg caagactgga cttgcaaaag ntccagcctg 420
tatggctgga gttcccatg cctgccaatc tctgtctgac tgcgagtcag ctccgatact 480
tcaccagatt cagccacctg ggggagctgg aagtgaatct cctcgtagct gagccttctg 540
atgagactgc agccccggct gacacctgga ttgca 575

```

<210> 65

<211> 834
 <212> DNA
 <213> Homo sapiens

<400> 65
 cagcagttgg tgactccttg gtgggagccg tgtcccaccc ggtcctgata ctgccgtctt 60
 ctctttcaca gtcctccagg cttggggccag ccttgggggc agcagagctt ctgggctgac 120
 atgggctcat tgctccttct ccaagccctc tgaggacatc aaaagcgtgg acgcatcact 180
 ttccaccatc ttgctgcccc ctgtccctcc atcctgaggc ctctaagca catgtgtggg 240
 gtggcaggca cactgctgat agctgtggat gcggccgtga catccttcac ccttgcccc 300
 atggcatgca tgatccatta gggaggaccg tctgcacaaa ggtctcttgc cctgtgcagc 360
 ttctctgcaga ctggacttgc aaagtccagc ctgtatggct ggagttccca tgcctgccaa 420
 tctcctgtcg actgcgagtc agctccgata cttcaccaga ttcagccacc tgggggagct 480
 ggaagtgaat ctctcgtag ctgagccttc tgatgagact gcagccccgg ctgacacctg 540
 gattgcagca ctcatgaaag accctgagca gcaggaccag tttggcagag cccgaattcc 600
 tgacccacag gaactgggag ataaaactct gtggttttaa tcttctcatt ttagagtgt 660
 cagtgtccat gtggtgtgaa cacgttcat tcaacctggg cccttgggag agatgtgtgag 720
 tggttcccg gctgtcccca ctccacacca tggcagtga gagctgtctga agtacatgct 780
 tcatagtccc ttgcgtctcc tctatgagta cagttcctgt ttgtggagta gcaa 834

<210> 66
 <211> 437
 <212> DNA
 <213> Homo sapiens

<400> 66
 cgagaaagaa aaggtatagc ttaaagtggc ttttgagcag gcatgagttt atggaaccaa 60
 ggattcctgt gaagacattt tcttttgata aaagaatatt gataagaata ttataccaaa 120
 ttgaacaaaa gtagccacag tatgaaggat tcagtacatg gccaaataac ttatttcaaa 180
 atagtttaga gttatattcc ttgaagacgg aggttggatg gggattaaat tttgtaaaga 240
 cgccaatggc tgtaaacaa aagagctgag atggatgtgc tcttgaatta aaaataaaaa 300
 tattttaaat atactattac atcataaaca ttctatgtct ctacttttcc atctagaagc 360
 aagaattctt tagtactttc cgagcatcta ctgtgtagac tatcttgtgt tatgaccaat 420
 tgcttatatt tatttac 437

<210> 67
 <211> 80
 <212> DNA
 <213> Homo sapiens

<400> 67
 acaaaacccat atgcttcaac acctcaggtt gaccatttgg ggggagtgtg tatgggtgtt 60
 ttaagatggc ggggtatgcc 80

<210> 68
 <211> 663
 <212> DNA
 <213> Homo sapiens

<400> 68
 gtgtagagca tggaagcagg gagaccagtt aggagtctat tgtaatagtc ctggtgagag 60
 accacagcgg cttggactaa gatggcaact aagataatga tgggtgcagg gcccctcttc 120
 aatggaggca ttgccagcct tctggccatg aaggagaaag tgatttcaac taaccaggga 180
 aactcttacc tctaaatgga gatacttctt gataacagaa gaaactgggc atctaaccga 240
 gaaataccag ctgagtagga gaagagaaaa ggcacagcc agtcaagggt tcagaaggct 300
 gccaacagtc tttgtaagcc accttgggag tagatgagaa cggcaatcaa tcaacatggt 360

27

```

ttggtgaaca aaccatatat tacaaagtgc ttctgtgaag tctgcatcct cacaactaat 420
gagttagaca tttctcattg tttctgctca cccaggaata ccatgctgtg ccagctcttg 480
ccatttatta accaactgat aatgggtgcag tgctgtagtc atggaagcta tttcaaaagg 540
ttaaggaagt ctactggaat cctggttctt ctagttagcca ttcagactta tttttaaagt 600
ctcattgaaa tgtaatgcat gttatggaaa gtcaggatga aataaaattg agattttttt 660
ttt

```

```

<210> 69
<211> 695
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (309)..(482)
<223> a, c, g or t

```

```

<400> 69
gaaacacaga aagagggggag aaacaggaga ggggaaagag agaggagaga gaaaccaagg 60
aatgtgaca tataataatt ttttaaagaa tattttttca tttttttatt gaggtataaa 120
atacatgtag taaggatagt caataactca aatcttatgt gattttttta tgtacatgta 180
tacctgtgta cacctgtgta accactacct aagtcaagat agagaacatt ttaatcatct 240
taaaagattt cctgtgtctc ttcccaccaa tacctgctga tgagcccact ctccttacag 300
ctatcagcnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnccttcatg ttaatgaaca tttgaattgt tttcatgttc ttgttatgaa tcaacatggt 540
tatgaatagt ttggttatga agagttttac acatgttttt agtctatttt gtttctctta 600
aatatatact tagtcacggg attactggtc atatagtata ggcaggcaga tgttcagctt 660
taattgacac aaccaactgt ttttgaaagg ggttg

```

```

<210> 70
<211> 739
<212> DNA
<213> Homo sapiens

```

```

<400> 70
ggtttctctt catggacatt gtttgcattc acatgtgaca cttaggaatg atctgttttag 60
tctcaatcac tcaactcctg atctgcctgt ctctctctga gataacaaag gccttaaatgt 120
ttagccacct gcatcagagt tggtagggtg gtttgaaaca attcatccta atataaaaag 180
aacagctttt gtaagggggc actgagtgtc tcaaacagcc gcatgggcag gaagagtgtc 240
cagtcacagt ttggttgaac ttgtcttggt gccctaaggc ctcctatgaa agactgacag 300
gcttggactg aatcttgtga tctggacacc aagggtcacc tgtgggcca gagctagctc 360
tgaagaatgg ggtagtcttct ttgagaacct ccacagcaaa agtttgggtc tctgttccca 420
atgcatgtcc cactttacca gctacatccc ccagtacctg cccatggctc atgactcatg 480
aaatataaaa ctgtagtagc aggcataaact ggttcagacc tgccagggtc atgtgggaac 540
tatcattggt acaaaaactc taagtgtgga gaagactgtg gtagacaaga ggggacatgt 600
ctgttctaaa cgcacatcag aaacttccaa tgactatggc caagttagat aagggtgtac 660
agaacttctc aggacatgca gacctatgtg tcaactcataa ctgaaattca aataaatatt 720
ttgtggattt cctgtggac

```

```

<210> 71
<211> 9883
<212> DNA
<213> Homo sapiens

```

<220>

<221> unsure

<222> (7153)

<223> a, c, g or t

<400> 71

```

ataagaataa aaattacccc aaatttccaa atcaagaagt aatcatgggt cagggtttggg 60
cagatgttct ttctaggcat gaacacacgt tatctcattg tttacttaac accgggttat 120
aaacattttac ccatagcatt tgaaaggtag ctatagatag aaaagaatca gagaagttct 180
aaaacagctc ttgcgctttg tttcaaattc tctgcaggaa agatgagggtc ttcagccttt 240
tttttagctg gacggcaccg ttgcagcagt ggtgaacagg gcaactggatt gagtcaggaa 300
accagctgt gaccttgggc aagccacttg cctctttga gcttcaactc tgctaaggca 360
agggggcgcta ttctgacctc gtctgcccac ctcacaggct ctggtgaagt ccttgatttg 420
aacgccttta gctcccaagg ttgtggtttg gagatagggtc aggtcacatg accatgaaga 480
ctgaaggaga aacgtggaag cacgtgtgcc tgttgccttc tttccaactt aaaatgcttg 540
gtgatctcct gaagactcca gcctcctcctc tgggaagcca ggatccacag accctttacc 600
tgcggtcat gggcagctccc agatggtccc cctcccaac agaggggggtg cagtgaagacc 660
tccggaagtt actgcctctg ttacctcaa agggattttc agatcagaca gccccctac 720
tccaagggac gtgtgtggag cttggtacct ttatttatct cctgctccaa cccctgtgga 780
ctgctgcac ccagaatggg gcctctcctg ctggcaagtg gctgagaacc tccactccac 840
tcagcagggc tgttccccat ttaccgaaaa gctccgagag aaaataacta accccatggc 900
gccgctgtag ctactggcag agcctcctgg tccccacctc tagcgctgt ggtttttgtt 960
tcatgcagag tgagcagtga atctgggac ccatcagcag tcagtttggg tgctgagcag 1020
gcacaatgat agatgttggt gaagggtagt tgtgaggata ttaattaata ttaacatgct 1080
agttatatta atattctcat tggaatttgt ggggctttgc aagttatatt tcaaatatat 1140
ggttttatgt aatcctttta actgcctga aaccttgaaa ttattgcacc tattttatag 1200
atggagagac tgaggctcag aggggtgaat tgcccaagat cctgggggag gaagcaccca 1260
ggttttctgg ttttgagtc cactctggga aggccagtggt attgattgc tgcttagacc 1320
gcccttgagg agcttgacg cacactggga aggccagtggt attgattgc tgcttagacc 1380
tggaagcac gtgaataaag cttcagggtta aaacctatgg ggttccagga ggcagcagtc 1440
ggctctgcct ggggggtgagc tgaggagccg gtgctctctg gaacaagggg agttgggctg 1500
aggctcagtg gacagtggag gttggcaggt gaagtgcagg aggtccttgc agggagtggtg 1560
accaccttga gcacacacag aggaatgaga caggcaggtt actcaaggag cagaggtctc 1620
gtgaccactt cccagagcat gtggggctcct agcctcatct ccaggaggag aaagtgcac 1680
tatacacaga tttgtcaatg gagtttaaat aggatgtggg aaaatctaga ttttccaaaa 1740
cagtacatat ttgctttgag aagaaaggta gatgcaggat gcatagggtta gataatttta 1800
atagcagtaa cctcagagca tgtaagtatg atttgattta ctggagtgc tgccgctctc 1860
agtcagtggg agcacggctt gggctgggag atgaggttga caagggttct ttctctcaaa 1920
tgcttctctt ggtttgctaa gagtatctc ctactcgcc gggggcagaa gacttttctc 1980
tcttttccca gtttgagta gttgggccag atttggtgaa gtgggagaaa ggccctgcct 2040
gcttctacat agagttggct gtctgactt gatactcggg gtgcttcca gagaccgcc 2100
tccatctcct caactccctg gcttgatgct taggtggtga tggctgttgg gcacaggagt 2160
tacataacag atctgtgatg gaccaggag cagagccagc tgagtgaatg tcatggagtg 2220
ggagtgggtc tgcatggctg tgggtgcccc tgcagcttgt gcagggtatg tggcaagagg 2280
tgctcaccac tcatctggaa tggctagact ggaagcactt gccctcttg ggctctgcac 2340
ccccacccc tcccactgg cctgctgct catcttcatg ggcgctggg gagaccaatt 2400
atggtgctt gtcatagtgg ctcaggtcac cgttcacact tctgggacc aggacatcag 2460
agccctgaga agggcaagg ggccaagtgg gcctagcctt ttactgacag ctgggaaatg 2520
caagcgtgtg gaccagagca ccaagtgagt tggggccggg gtgggttcag caccgtgtcc 2580
ctaccagag ctccatttgt tgaaaacagc ctttctctac cgtttcttca cttggacaac 2640
tttaaaactat gtattggctg gtcgcggtgg ctcacgcctg taatccagc actttgggag 2700
gccgaggtgg gcaggttaact tgaggtcagg aggtcgagac cagcctggcc aacctgtaac 2820
aacctcatct ctactaaaaa taaaaaatt agcctcagc ggtgacacgc acctgtaac 2880
ccagctactc gggaggctga ggcagaagaa tcgcttgaa ttgggaggca aagattgcag 2940
tgagctgaga ttgcatcacc gcacttcagc ctgggagaca gagcgagact gcatctcaaa 3000
aaacaacag aaacctacat attttctata tttcccccac cattgaggct catttcttgg 3060
atgaacaatt taaatgtact gtgctctctc ggcaaatatt tccaaaatta cagatgttcc 3120
tatactttca ccggcagctc tgctcccg aattttattc acggatgggt taacacgtgt 3180
gcaaaatgat ttatttgcaa ggttcgtcat tgttgcttta tttttaatag caaaagattg

```


gaggcagctt	aaatgttcat	tgcagggg	caatgaacaa	accatggccc	gtctaaacat	3240
gggataccgc	gtggccataa	tacataagat	ggacgctcaa	cgcactgtgc	cggattgagc	3300
agcaaggtgg	attgccgagg	gaagaagcag	gtctgggccc	tgtgtctcgg	agctgccatc	3360
agtgtaaaag	ggaagagaat	caaaagtgtc	tttgcttgtc	tatgcccagg	gggtctctgg	3420
gcagacaccg	caagtcggtg	attgtgatgc	ctctggaggg	gggtgctggc	atgggagatt	3480
gcttgtttgc	tggagatccc	atgtaccttt	tgattgctga	agcaggtgaa	tgtacgcctt	3540
tccaagaaat	taaaatgggc	caggtgcggt	ggctcacgcc	tgtaatccca	gcagtttggg	3600
aggctgatgt	tggaggatca	cttgagggtga	ggggttcgag	accagcctgg	ccaacatgat	3660
gaaaccccat	ctccactaaa	aatacacaaa	ttagccagac	atgggtggtac	atgcctgtaa	3720
tcccagctgc	tctagaggct	gaggcaggag	aatcatgtga	accctagagg	ctgagtttac	3780
agtgagccaa	gatcatgcc	ttgtactcca	gacctgagcta	cagagcgaga	ctctgtctca	3840
aataataaaa	taaaataaat	taaaaacata	aggactgtaa	ccttgccctcc	tgcccagttg	3900
aggaaggtca	aggttctggc	tacttctcaa	gtacaggagc	ctcactcagg	ccccagacca	3960
ctaatacaaaa	aatatgtgct	tggttctcac	aaaggggccc	agtgtgaggg	cttgggtggt	4020
gcttggtaaa	tacgaccccc	gggtcccgcc	ttggagagat	ggagccctct	ctggggccct	4080
tggacacact	gctgttggct	gactttgtca	ttttcaacct	ttgctccgat	tggtcacgt	4140
catgatttct	gaaacctttg	ggggcttccc	cactgacaga	aagatacact	ttaactcagc	4200
actgggcac	ccaggccctc	tttactgggc	ctcttcttga	gccgcacttg	gctgtcacc	4260
ccttctctct	tctgcccctc	taactcccca	cctccgtgcc	tttgctcata	tagttccctt	4320
tgcttgcctt	tccgtccaga	gcagtctcca	cgtgcccagg	tcctgtctga	ctttcaaggg	4380
ccagcttagt	ttccacttct	gcactgcctt	ctgacctccc	tggttctctg	gtaaaactgcc	4440
cagatcaagc	cacacaatgg	ttcctgcacc	caaggaagct	ccctggggcc	ccctcctggc	4500
cactcgctct	tgcgaggtag	tcaccactca	caccttggca	ctttcgcgtg	gtgcttggcg	4560
ctgctgtttt	gggcctccca	cacacagagt	gtacagaacg	gactcctcgg	tgtctggctg	4620
ccttcccgc	gcactgtcag	atcatccagg	ttgctgttag	tggccctttg	gttttttct	4680
ctgctgcgta	ggagttcacc	aaatatacca	ctattttattc	atttctcctgt	ggacaggcat	4740
tgggttatgt	ccagcctctt	cgggtgaattc	attcttgtct	ttggggggcg	gtgtgcgctc	4800
tttgctgggt	atacaccag	ggtgggttga	tggttacct	gactcagaat	gtgtttgcat	4860
gaatgaaatt	cagggttggt	tgagaaatct	aggggtgtcct	ggctggagcc	aggcttcttg	4920
attacaggga	cagagcagg	acagggatcc	tggtttagac	agcctgtctc	catggggtgg	4980
tagcattgtt	ggggtgcagg	atgctgaatc	tgcaggggac	ctatccgctc	agtgccag	5040
gggattttag	ctggctggaa	aggtggtcac	atgtagagg	gctcaacaat	ccagctaaag	5100
aggctgagcg	ttgggtccatt	gttctcaatt	tgagagaaaa	ctgagatcat	caaaattagg	5160
actggtatgt	actaaaggaa	agaaccta	tacaaggctg	aattgagtaa	gcctcgtg	5220
agggactttg	gatttctttg	ttgttccctt	ttatttctgc	acccccacc	aagtgcaga	5280
tatgtacatg	attggatgat	tttgctttcc	tggttgagag	attcctggga	acttgccca	5340
ggagaaaggg	gagaaatgtg	gagccgctag	agtggcctcc	gcttgtttgt	gttgattgaa	5400
ggggagagcg	aaggagagct	gtggaccctt	gaccccttgt	gagggcatgt	gatccttttc	5460
aaaaggctca	ccaggcagaa	gtgcttgccc	aggggcccgt	ctttccctct	aatccccctc	5520
ggagaagggc	caggctgtgg	gttgcgtgac	tgtctgatg	tggatcagcc	tcccccaata	5580
atgcagctgc	ccagaagctc	agagagccca	ggcaaccccc	aaaggcagga	gggcccgtg	5640
tcattcccgt	tgtcattccc	aggcggtgg	agtgggagca	gagcggtcag	ttcagatgaa	5700
cagtgtctga	gtctgacccc	aaccagcgag	ttatggtaag	atggaagggt	ctccatctat	5760
attaataaag	agaacaaaag	ccctcccagg	ctgcatgaat	attccaggga	tatatatgtg	5820
aacgggttgc	cagtttagct	tggcctgtgg	gtggcagccg	cctgagttag	cacttcgtgg	5880
ctgcagctct	aaagggtttg	gatctgaaac	taatgaatga	aaatatgacc	tcagaagatt	5940
taaagagagc	aaataccag	caacagaacc	tgggtcccag	agactgttgg	gagcatgaaa	6000
tcccaggctg	gccgaaggag	gaagtgggag	agcaatggca	gctgacatca	catggtgcca	6060
gaccttctca	gtgctttctg	tggttactca	ttattccgtc	cctctctctc	agaggcagg	6120
atggctgctt	ccccatttta	tagatgagga	agctaaggca	aggagagggt	gtgtaacttg	6180
ctcacagaca	caaagctagc	cagtggcaaa	gctggagggtc	aggtctaggt	ggtcaggctc	6240
cagagtctctg	cggatttctc	agcacggcag	tggcagtcgg	aagaaccatt	tgtcagggtga	6300
ttgtgggcaa	atgacgtcag	cccttcaaac	ctctgttttg	catctgcaag	ctgcttgctg	6360
ctgcaacaaa	ttaccagaaa	cttagtgact	taaaacacaa	attaggtcgg	gtgctgggtg	6420
tcacatctgt	aatcccagca	ctttgggagg	ctgaggtgag	tggatcactt	gaggtcagga	6480
gttcgagacc	agcctggcca	acatgatgaa	accctgtctc	taacaaaaat	ataaaaaatt	6540
agccaggcat	ttggccgggt	gtggtggatc	acgcctgtaa	tcccagaact	ttgggaggac	6600
aagggtggcg	gaacacaagg	tcaggagtgc	aagaccagcc	tgaccaatat	gggtgaagcc	6660
tgtctctact	aagaatacaa	aattagcagg	acgtggtggc	acgcgcctgt	agtcccagtt	6720

actgggagggc	ggaggttgca	gtgagccaag	atcacgccac	tgactccag	cctgggtgac	6780
agagtgaagac	tccatctcaa	aaaaaaaaaa	aaaaagtaca	aaagagcaaa	acaaaacaaa	6840
agttatgaaa	atgaaaacct	gagccatcct	ttatcttatt	tccccaatc	cactaattat	6900
taacagaaag	taaaagctat	gaaaaatgaa	tgaaagtga	tgcaatttcc	ttgaagtgtg	6960
ttagaacctg	cctttagtgt	cagctatggg	ttccctcatg	aaggtcagct	gagccatgac	7020
ccatgaacca	tggaagcttg	actctagatt	gaccatcttg	agatgccaaa	gatgtccacg	7080
tcctaatccc	atgtgggaga	cagaataatg	gccctgcaga	ccttcccagc	tggccatgac	7140
ccctcatttg	acnagctctt	cccttctctc	tgaccagcac	catgttcttc	ctggtgacaa	7200
gccttctgct	ctgtgagtta	ccacacccag	cattcctcct	gatcccagag	aaatcggatc	7260
tgcgaaacagt	ggcaccagcc	tctagtctca	atgtgagggt	tgactccagg	acgatgaatt	7320
taagctggga	ctgccaagaa	aacacaacct	tcagcaagtg	tttcttaact	gacaagaaga	7380
acagagtcgt	ggaaccaggg	ctcagtaaca	acgaatgttc	gtgcacattt	cgtgaaattt	7440
gtctgcatga	aggagtcaca	tttgagggttc	acgtgaatac	tagtcaaaga	ggatttcaac	7500
agaaactgct	ttatccaaat	tcaggaaggg	aggggtaccgc	tgctcagaat	ttctcctgtt	7560
tcctctacaa	tgcggtttta	atgaactgta	cctgggagag	gggtccgacg	gccccccgtg	7620
acgtccagta	ttttttgtac	atacgaaact	caaagagaag	gagggagatc	cggtgtcctt	7680
attacataca	agactcagga	acccatgtgg	gatgtcacct	ggataacctg	tcaggattaa	7740
cgtctcgcaa	ttactttctg	gttaacggaa	ccagccgaga	aattggcatc	caattctttg	7800
attcactttt	ggacacaaag	aaaatagaac	gattcaacct	tcccagcaat	gtcaccgtac	7860
gttgcaacac	gacgcactgc	ctcgtacggt	ggaaacagcc	caggacctat	cagaagctgt	7920
cgtacctgga	ctttcagtag	cagctggacg	tccacagaaa	gaatacccag	cctggcacgg	7980
aaaaacctact	gattaatggt	tctgggtgatt	tggaataatg	atacaacttt	ccaagctctg	8040
agcccagagc	aaaacacagt	gtgaagatca	gagctgcaga	cgtccgcac	ttgaattgga	8100
gctcctggag	tgaagccatt	gaatttggtt	ctgacgacgg	gaacctcggc	tctgtgtaca	8160
tttatgtgct	cctaactcgtg	ggaacccttg	tctgtggcat	cgtcctcggc	ttcctcttta	8220
aaaggttcct	taggatacag	cggctgttcc	cgccagttcc	acagatcaaa	gacaaactga	8280
atgataacca	tgaggtggaa	gacgagatca	tctgggagga	attcacccca	gaggaaggga	8340
aaggctaccg	cgaagaggtc	ttgaccgtga	aggaaattac	ctgagaccca	gaggggtgtg	8400
gaatggcatg	gacatctccg	cctccgcgac	acgggggaac	tgttttcttg	atgatgctgt	8460
gaacctttat	atcattttct	atgtttttat	ttaaaaacat	gacatttggg	gccaggcgcg	8520
gtggctcacg	cctgtaatcc	cagcactttg	ggaggccaag	gcaggcggat	cacttgaggt	8580
caggagtctg	agaccagcct	gcccacatg	gtgaaacccc	atctctacta	aaaatacaaa	8640
aaaattagcc	ggcggtggg	gtgggcgcct	atagtcccag	ctacttggga	ggctgaggca	8700
ggagaattgc	ttgaacctg	ggaagtggag	gttgagtcga	gccgagattt	gtgccactgc	8760
actcccagcc	tggcgacag	agccagactc	catctggctc	aaacaaacag	acaaaacaaa	8820
acaaaataaa	ataggcccag	tatgatggct	catgcctata	atcccagcac	tttgggaggc	8880
aaggcagggtg	gatcacttga	ggtccggagt	tcgagacaag	cctgggtcaat	acagtgaaac	8940
cctgtctcta	ctaaaaatac	aaaaattagc	tgggcatggt	ggtgcatgcc	tgtaacccca	9000
gctactcggg	aggctgaggc	aggagactca	cttgaaccgg	ggagatggag	gttgacgtga	9060
gctgagattt	gccactgcac	tccagcctgg	gcgacaccgt	gagactccat	ctaaaaataga	9120
agaaaagggtt	tctcttcacg	gacattgttt	gcatctacat	gtgacactta	ggaatgatct	9180
gtttagtctc	aatcactcac	tcctggatct	gcctgtctct	ctctgagata	acaaaggcct	9240
taatgtttag	ccacctgcat	cagagttggg	gaggtggttt	gaaacaattc	atcctaatat	9300
aaaaagaaca	gcttttgtaa	gggggcactg	agtgtctcaa	acagccgcat	gggcagggaag	9360
agtgtctcagt	ccagttttgg	ttgaatttgt	cttggtgccc	taaggcctcc	tatgaaagac	9420
tgacaggctt	ggactgaatc	ttgtgatctg	gacaccaagg	gtcacctgtg	ggcccagagc	9480
tagctctgaa	gaatggggta	gtttctttga	gaacctccac	agcaaaagtt	tggctcctctg	9540
ttcccaatgc	atgtcccact	ttaccagcta	catccccag	tacctgcccc	tggctcatga	9600
ctcatgaaat	ataaaactca	gtaggcaggg	ataactggtt	cagacctgcc	agggctatgt	9660
gggaactatc	attggtacaa	aaactctaag	tgtggagaag	actgtggtag	acaagagggg	9720
acatgtctgt	tctaaacgca	catcagaaac	ttccaatgac	tatggccaag	tgagataagg	9780
gtgtacagaa	cttctcagga	catgcagacc	tatgtgtcac	tcataactga	aattccaata	9840
aataattttgt	ggattttccaa	aaaaaaaaaa	aaaaaaggcg	gcc		9883

<210> 72

<211> 93

<212> DNA

<213> Homo sapiens

<400> 72

gtttatattaa aacaatagaa acattaatct gtctgtcttt totccattct atccattcgt 60
 tctttaatgt ggtcactttt gaatgctgta tac 93

<210> 73

<211> 299

<212> DNA

<213> Homo sapiens

<400> 73

ctcgagcgt cacatattac cacctctgta aatccttttc taacttattc agggtgaccg 60
 aattctgtgt ttctgtgccc ccttaatact tgttatataa gtctccttcc ccaaccaccc 120
 ccacacttac cacatcacgt tagcaagaat gagagcaatt tgagggcagt ggctttgtat 180
 cttattttata gccctggcac caaacagtt tgtaaaaagt taatctggtg cagggtggca 240
 taacacataa gagtctgttt cttttgagat atttggcaga gggtgtggtg tgcggagat 299

<210> 74

<211> 94

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (85)

<223> a, c, g or t

<400> 74

gctgtgttta tgctgctggc tgtactggga ggaatatggt cctttgtctc tgaccagga 60
 gtttcatgtc ttctgccaag atacnttaca tgga 94

<210> 75

<211> 433

<212> DNA

<213> Homo sapiens

<400> 75

gctgtgttta tgctgctggc tgtactggga ggaatatggt cctttgtctc tgaccagga 60
 gtttcatgtc ttctgccaag atacattaca tggatagata cattaggtag gtagatacat 120
 tagatataga tagatacatt agatatagat agatacatta gatatagata gatacattag 180
 atatagatac attagatata gatggataga cagatagata cacagataga tagatagata 240
 gatagataga tagatagata gatagataga tagattcatt tatttattga gacagagtct 300
 tgctctgtca ccgaagctgg agggtagtgg cttgttcttg gctcactgca acctccacct 360
 cctgggttca ggtgattctc ctgcctcagc ctccacagca gctgggatta catgccacc 420
 tattttgtac ttc 433

<210> 76

<211> 334

<212> DNA

<213> Homo sapiens

<400> 76

gctcgagggt aatggaccat tcgggttata tggttcatat tttttgtca tttttatgtc 60
 atgggtgttta tcttttctgt gctgatttgt aaaagctatt ttaaaacct tcatctgcca 120
 tatatgttac atttctttcc tgctttctgc caccttccaa tttgttacca actttcttct 180

32

```

ccaaccttgg gccactggca tataactca ttttaaataat cagaacttgt agtgcctctt 240
gaaatgcaga cagactatgg ttcattctgc aactgcatat tagttaacag gcaaaaatac 300
cttagtaaga gaaagtgtct tttccttcta atgt                                     334

```

<210> 77
 <211> 547
 <212> DNA
 <213> Homo sapiens

```

<400> 77
ggcttatatg tggagaactg acgtctgaac ccagatctga ttcccaagtg taatactttc 60
caataggcag ccttatatct ctgtacctca aaagagaagg ctatattatt taaaagatta 120
ggaattgtcc tatatggttt taaaatacac ttgctatagc acaataataa gtgggttagt 180
ggtgactgct actcctgtga gtttggttta aaaacagccc agtttgtacc ctgttggtca 240
tgataaaagc ataccacct tactttgaga attttaacca tagagcacia tatgtgtcaa 300
acaagctaaa aaagtattct tttcagttgc attttgatgg acattgaaat tgcttagact 360
ctttgaccaa aagtacaaac tgctgttaaa ctggtgacaa aatctgtttt catggacgct 420
aggctactta agctttatct tctcctaag cattctctgc ctttgtaaag cactctagca 480
gcagtatttg cttagcttct aattttggtt ttgcttttgc gttttctctc tttctcttgg 540
ttgttcc                                     547

```

<210> 78
 <211> 263
 <212> DNA
 <213> Homo sapiens

```

<400> 78
tcgaggggtg aaatgagtgt cattagccaa gtgacattta agtgccttgg tttgtctgct 60
tgcttttctg tggattgaaa aaaactgacc actgttaata tgattgtaca gtgacactgg 120
aaattatgag atgtgtgtct ggtagtctct gcttgatatt cagttgagat gcataccaag 180
tctgataatg cagagctttt ccatttcatg tgtctgttta ccattttcat gatcttaagc 240
aataaacatt tcttgacaac agc                                     263

```

<210> 79
 <211> 765
 <212> DNA
 <213> Homo sapiens

```

<400> 79
gcgggaagag cacgcagccc tgcgagtact atttccgcgt gtaccactcg ctgtgcccc 60
tcagctgggt gagtcggcag agggggcgcc gggccaggcg tgtgcagggc tcggccgagg 120
ctgagccggc gtcccgtctc ctgcctttct gcttcccagg tggagagctg gaacgagcag 180
atcaagaacg ggattttcgc cggcaaaatc tgactgcccc agcgcgggctt cctctgaaga 240
tgcaagtatc ctgcattctt ttgtctcgcg gagccccggg tctcgggttat ccaccttac 300
ctcccagtg ctaagccacg aataatgcca ccagccttcg agttccttgc ttcccttgc 360
ctggctctca cgtgtatgat ggggttctca ggcccaggct tcgaccagag gacctctgc 420
caccaccgtt tcttctgtc cttgagctac cttggtgaac tcatgacccc agggccctgc 480
tccaccagga tgtcccccag gtcctgccag ctgggaagtg ccagcatgaa cgccctcaac 540
ttcgtggaag ccagggtccc ctgcagctga gggacgcaa gcagacacac ctgccctccc 600
cagccagctc ctgtctgtat gggcgagatg actgagagcg cccacgtccc taaggctgtc 660
ctgacctcc atgctgcgac aaggacaggg aatggtcgg cactatgggc ctggtgtctc 720
ccctccccc ccaccgggtg ctgcccagct caagccagaa gtgac                                     765

```

<210> 80
 <211> 162

<212> DNA

<213> Homo sapiens

<400> 80

```

cgctgcctca agaccaggac ccgcccggg aagagcacgc agccctgcga gtactatttc 60
cgcgtgtacc actcgtgtg ccccatcagc tgggtggaga gctggaacga gcagatcgaa 120
gaacgggatt ttctgcctgt gcaaaccatct tgacttggcc ca 162

```

<210> 81

<211> 986

<212> DNA

<213> Homo sapiens

<400> 81

```

agcgggcggt gcacgacggc tcccattggc tggggctcgg gcgtcctagc caatccggcc 60
gcgggggtgcg tttctcctga cccgggtggg accgcacccc gcggactcag aagegagcgg 120
caccgccgga ccatcccaca gcagatccag tggccgccc aa cgtcaggctg gagttgcctc 180
cttcgtggat gttggatgtg gaagcccagg agccccccaa ggggaaatgg tcgacgcgcg 240
ccttcgaccc gcgcttcccc agccagaacc agatccgtaa ctgctaccag aacttctctg 300
actaccaccg ctgcctcaag accaggaccc gccgcgggaa gagcacgcag ccctgcgagt 360
actatttctc gcgtgtacca ctgcgtgtgc cccatcagct ggggtggagag ctggaacgag 420
cagatcaaga acgggatttt cgccggcaaa atctgactgc cccagcgcgg ctctctctga 480
agatgcagtg atcctgcacg tttttgtctc gccgagcccc ggggtctcgg taccaccccc 540
tacctcccag tgtctaagcc acgaataatg ccaccagcct tcgagttcct tgtttgccct 600
tgctcgtggt ctccacgtgt atgatgggg tctcaggccc aggccttcgac cagaggagcc 660
ctctggccac caccgtttct tctgtgcct tgagctacct tgggtgaactc atgaccccag 720
gccccctgct ccaccaggat gtcccccagg gtccctgccag ctgggaagtg ccagcatgaa 780
cgccctcaac ttcgtggaag ccaggctccc tgcagctgag ggacgccaa cagacacacc 840
tgccctcccc agacagctcc tgtctgtatg ggcgagatga ctgagagcgc ccacgtccct 900
aaggctgtcc tgacctccat gctgcgacaa ggacagggaa tggtcgggtca ctatgggcct 960
ggtgtctccc ctcccccatc aaccgg 986

```

<210> 82

<211> 369

<212> DNA

<213> Homo sapiens

<400> 82

```

aaccctaagat gactcgtctt ttgggtgggag aattcactct gttcatgttt catttaacaa 60
ttgatctact gtacttaatt acctttggct tattttacat ttattggttt atcttggtt 120
tttcttccct ctgatctggt tategatttc ctttttcttc ccctgttgca ctttccattt 180
cattattggc agctgtccct tctctgggg tctaatcaa acacatatc tttagcacat 240
gcctcgatgg ggattctttt cgcagcacc tcacttggag ctcacagaac ctgtcactct 300
gtaggttctg gtcttttttc agcttaggaa catctatttg ttgcttgatt tgattattgt 360
tagtttggt 369

```

<210> 83

<211> 923

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (354) .. (565)

<223> a, c, g or t

<400> 83

```

aacccaagat gactcgtctt ttggtgggag aattcactct gttcatgttt catttaacaa 60
ttgatctact gtacttaatt acctttggct tattttacat ttattgggtt atcttgtgtt 120
tttcttccct ctgatctggt tategatttc ctttttcttc ccctgttgca ctttccattt 180
cattattggc agctgtccct tctctggggt tcctaataca acacatatte ttttagcacat 240
gcctcgatgg ggattctttt cgcagcacc ccatctggag ctacacagaac ctgtcactct 300
gtagggtctg gtcttttttc agcttaggaa catctatttg ttgcttgatt tgannnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnccctg gataggaagg gataggaaga gactacttgg 600
tgccatgggg taggggtgag ggtataagta gatcagagtg ggaagacctc agccttgggt 660
ggcttgtctc tgcttcttgc cagggtgggag ggctgtcca cacctggatc cccgtaccac 720
agtgccagcc atgcccttcc ctgggctacc attgtccctt tcctcaccca gttggtagag 780
gagtcaggag gtgggaggcc gtgggctttg gttttataat gtaaccactg tgggggtggg 840
ggaggatggt gaaccatgta tttcagtga atatttaata tatttaata tcaataaaat 900
caaacctctt gtaaaaaaag ccg

```

<210> 84

<211> 338

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (12)

<223> a, c, g or t

<400> 84

```

ataatttttt tntttttaaa ggaaatgaac gtggaggact ggggtgaagg gccagcctgg 60
gtagttaaat ctttttggga agacatgact ttaaggagat tccctgcttt gtgacagggt 120
gctccatgct gtcttgggga caagggcctg tactgccttc aaatctgggc tcaccccaca 180
ttttgtgag gggaagatag ggtgggggga taaggaggag aaaagactct agcttttttt 240
ttctatgcat gatatactgt gtgggtttat caagagtgtg gacacagttg ctgttctcaa 300
ataataggcc aaataaaatg cgattctttt tttctttg

```

<210> 85

<211> 436

<212> DNA

<213> Homo sapiens

<400> 85

```

ataatttttt tctttttaaa ggaaatgaac gtggaggact ggggtgaagg gccagcctgg 60
gtagttaaat ctttttggga agacatgact ttaaggagat tccctgcttt gtgacagggt 120
gctccatgct gtcttgggga caagggcctg tactgccttc aaatctgggc tcaccccaca 180
ttttgtgag gggaagatag ggtgggggga taaggaggag aaaagactct agcttttttt 240
ttctatgcat gatatactgt gtgggtttat caagagtgtg gacacagttg ctgttctcaa 300
ataataggcc aaataaaatg cgattctttt tttctttgaa acacacagaa cagcccagct 360
ataaaacagg caactgagga agaaccacac cgcataccgg caagactcta gcatgtcaag 420
gtcaaagact ctccag

```

<210> 86

<211> 462

<212> DNA

<213> Homo sapiens

35

<400> 86
 aggggaacgtt ggatgtagtc aactgctgt tgggtgttact tagaccttca tttttccacc 60
 agactgtagt gttcaaaatt ctttttagta agagaaccct ttttttctga actttttaca 120
 accatctcca aattataaaa cataagactt ttttttagta aaaatatatt tttttacaag 180
 cacagtggct tgcacccatgg aggggagagg aggtgttttg tccttgagc tgetggcctg 240
 agagaacctt gtcacgtgg gagctgggcc attectacac agtggctctg caatgaccg 300
 gtgggtgggg aggcctgtga gtgggcactg gtaatgggaa cagctgtaaa accctggagg 360
 ccagcccg gagagtgacc ttaccagga aagttctggg aaacaaacca cagggaggct 420
 ttacaggaat ttttggttgt gccacaggc aaggcacatg ag 462

<210> 87
 <211> 1435
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (1012)..(1119)
 <223> a, c, g or t

<400> 87
 ttagaggtag aagaactgac tataagcaga agtgtttgag gaggtgcat ggagaacaag 60
 gggcatcatc ttggcccttg gcaggttggc aggatttgac ttggtgaaga gaacgagaaa 120
 ggggaacttta actgggagga ctactctggc ttgtatttct ccatcatgag gagattggc 180
 tttggaagtt gtagcttcca gagacctcg atgtttgcta acatgtccaa gctctacatt 240
 tattgattgt tggttctgtt catggctatg ttcaaattct tgtacctttt tgtctccac 300
 agtttcttgt ctcacccctg tcttccacct ctgctccccg ctcttgtctg gtctaattaa 360
 ctctctctgt tggagcagct tcccctcttg ggtaaactca gacatgaccg cagcaaagca 420
 gcgtggaatc ttctgtttgg tcagtgttcc cccagcttc cccgcagata cagctgcatt 480
 ggagccctg aagacaaacc agagaagtgc tgcacctgg ggggcaggag gctttgcttt 540
 gccagggct gggctcctga atgaatttg gtgcagcctt aacggccgag ttgtgctgtt 600
 gaagtgacac tgctctgtgt ccaggcactt catggagggg agaggagggtg ttttgtcctt 660
 ggagctgctg gcctgagaga acctgtcat cgtgggagct gggccattcc tacacagtgg 720
 tctggcaatg acccggtggt ggtggaggcc tgtgagtggg cactggtaat gggaaacagct 780
 gtaaaaccct ggaggccagc ccaggagag tgaccttacc aggaaagtcc tgggaaacaa 840
 accacagga ggctttacag gaatttttg ttgtgcccac aggcaaggca catgaggaaa 900
 agaaatgtaa ttatagtttg taagtcgatg aaaagaggca atgagtgaca tgaaatagct 960
 gctctaagtt tcttcttctt gtcggacagg aagaaatggg gttttatgca tnnnnnnnnn 1020
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1140
 taagcagata actgtataaa tgcataatta cacagcatgg tgagtgtctt gaaggataag 1200
 tgtggggagc ctcatctaga ttggaggatt gtgaaagtca agagacagga gagtcaagg 1260
 gaggaagggt gagtaagagc tatccaggca aagactgctt ggtagggggag tgtcccagca 1320
 acgggaaaca acctggaaaa aatatgacac cttaggggaa ctaaaagcag ttgtatgtgg 1380
 ctgatgcaca gacagggaag ggcagggaag gtgctgaaag aaggcaggag gagaa 1435

<210> 88
 <211> 459
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (437)
 <223> a, c, g or t

<400> 88

36

```

gtctggtttg agtctaggat gaaggtacct tcctccagga aggccctggt gttccttctg 60
ccagactcct gaggggtctcg ccagttcaag cccacttgaa gccagctcg tttggggtta 120
cttgaaccat ctgggggatt ccaactagta tcttttagctc ctgacatgag ctgttctact 180
gtgggctcag cccttgtctg agactgtatc cctatagggt cccggtcttc tgttgacccc 240
tcaccttctg tgggcctggg catggacctc tgatccttcc atctgaagaa gtgtcaaaat 300
aaaagtccat gcttccggga atcaggaagt cgcctcaagg caaaagtagc tgagtgtttc 360
tatactctgtt ttgttttctt ttctaacttc tctttttggt gggtaattct tcaccatctt 420
gttgattctt taagtctag cataacacac attttaaaa 459

```

<210> 89
 <211> 1263
 <212> DNA
 <213> Homo sapiens

```

<400> 89
gtctggtttg agtctaggat gaaggtacct tcctccagga aggccctggt gttccttctg 60
ccagactcct gaggggtctcg ccagttcaag cccacttgaa gccagctcg tttggggtta 120
cttgaaccat ctgggggatt ccaactagta tcttttagctc ctgacatgag ctgttctact 180
gtgggctcag cccttgtctg agactgtatc cctatagggt cccggtcttc tgttgacccc 240
tcaccttctg tgggcctggg gcatggacct ctgatccttc catctgaaga agctgtcaaa 300
ataaaagtc atgcttccgg gaatcaggaa gtcgcctcaa ggcaaaagta gctgagtgtt 360
tctatatctg ttttggtttc ctttctatct tctctttttg gtgggtaatt cttcaccatc 420
ttgttgattc ttttaagtct agcataacac acattttaaa aatccagttg ttttagttgc 480
tttctgtctc catagaaggt caccatgggt ctgagccctg tcggacctgg agcctggtac 540
catgaccagg gacagggagt cctcatgccg ttttaagcag tggatgata agttttatct 600
cttaggtgag tcaaggtcgg aaaagcttga gaccctgct ctaggggctg tacctgtccc 660
tttctccctt ttctcctgtc tggactaggg ttggaagggg ctgggtgggc atgtggagac 720
caagtagctg acaatcccca ggacctgtgg gctcagacac agggccctgc acctctcagc 780
ccttccggtc tcagctcagc acctcccttg cctggcccct ctttcttgcg tgagctccct 840
gcctctgcca ggaggaacct ctgtcctgtt tctagatgcg ccatactctc tcccacctcc 900
tgctctttcc tccagttgtg tgctctgtaa cctcttctc cctccaagga taaatcaaac 960
cctacctctt tatacaggag gaagtaattt ctgggttgat gtatgcatcc ggcagattca 1020
tgctgagcca acaggttagg ggctggagaa acagtgatga gcttaaccag gcctgcccag 1080
cctgcccacc ccgagtctgg tgagggtagc aaaaaacata aagtgggaatt gataaataat 1140
ataatctatc catatccata tttttatttt ttattatttt gggacgaagt cttgctctgt 1200
cactccagcc tgagctacag agtgagaccg tgtctcaaaa aaaaataaga aaaaaaaaaa 1260
aaa 1263

```

<210> 90
 <211> 554
 <212> DNA
 <213> Homo sapiens

```

<400> 90
gctcgagctg ttttcttcag gtgagtagaa caatggcatt ttaaactctaa gaggcacctc 60
gtaaatacat ttatttcaat tcctttccta cataggggaa gaaacagagg ctgcaaaaga 120
tttagttagt tcaagaaaaa acagtataat ttggagtttt tgactttgtg agttttgtta 180
cggcgctgac attcattctt ttgtgcttcc agtgatttca aatcttcaaa tctagagcac 240
attgtatgct gggcagaagg cacagtactt gaggattcag tggacagtga tacagaaaag 300
gctgctgtcc ttgggcactg atgagcctcg ggctactaca agtaagcagg cagtggcagt 360
aggtggaatg agggctgcag gtccctggcat catggatacc aatttggtgct tagaatggaa 420
gcggaggctt ccttgaagaa cagcggctca agctgagact tgtaggaata gtggttaatta 480
acaaacagac aggaagaaga gctttccagg aagacagcaa aacataggca aaggtctgga 540
gaggagagag agca 554

```

<210> 91

<211> 435
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (406)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (411)..(412)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (421)
 <223> a, c, g or t

<400> 91
 tattagtcca taaaggctat ttctagtatt aaacaatgct taagaatagc ttggatccat 60
 gaaaactttt gagaaggagg acaaagcaga cggaacctaa tctctgaaca atttcaatta 120
 catcttttac aagtggctgt tggctagtca ttaaaaaatga gccattcaca cttgtggaca 180
 ccttttttgc catgcagact tgacttgcaa agcctttatt atccctgggtt aagaacagca 240
 cagctaataa aaacgaatca tatggcttta aactacttgc atccaacagg gacatcctaa 300
 aaatgggtccg gatagtgact tcatgaccat ttaggctgca agtgccatag ttactaatga 360
 gaacagatat ttccaaatgg cggcaataga ttatggaaaa ttgagnaagg nnagagagta 420
 ntttactttc agcta 435

<210> 92
 <211> 580
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (551)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (556)..(557)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (566)
 <223> a, c, g or t

<400> 92
 aaaaaaactg tttagaaaac cttcatattt actctcccgt tcaaactatt ggccctgatt 60
 tttaacagata atcaaaagtc aggctgccaa acttattttc tttgaatttg gaatatcttt 120
 taaaatttgc ctttttcttt cttattatta gtccataaag gctatttcta gtattaaaca 180
 atgcttaaga atagcttgga tccatgaaaa cttttgagaa ggaggacaaa gcagacggaa 240
 cctaactctt gaacaatttc aattacatct ttacaagtg gctgttggct agtcattaaa 300
 aatgagccat tcacacttgt ggacaccttt tttgccatgc agacttgact tgcaaagcct 360
 ttattatccc tggtaagaa cagcacagct aataaaaacg aatcatatgg ctttaaacta 420
 cttgcatcca acagggacat cctaaaaatg gtccggatag tgacttcatg accatttagg 480

ctgcaagtgc catagttact aatgagaaca gatatttcca aatggcggca atagattatg 540
gaaaatggag naaggnnaga gagtanttta ctttcagcta 580

<210> 93
<211> 724
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (297)..(602)
<223> a, c, g or t

<400> 93
tactgatgtg cttttgattt gtctggaggg tgactactac ctctttgagg tgcctcctgg 60
gacctcaaaa atattaactt ttatactctg ttagcctgtg actttaagcc agaacattca 120
aagtacactg aagaaatgtg ttgaaaatct atgcaaccat ttctgcatta tgtactagca 180
aataaacaat ctttaatttc tggaattttc cattttcctc agtgatattg ttgattgatt 240
tgtagttttc tttctttgct aggtttcagt atcagggtcg taccaatttt tttctttnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600
nntgtgccat ctttatgaag tgaattatga agctttccaa tcttttttat tttgtagaac 660
agtttaaata cacaacaata tactaagttc ttagattgaa gctgttttta aatcacaaag 720
acag 724

<210> 94
<211> 586
<212> DNA
<213> Homo sapiens

<400> 94
ctaagacagt ggccaatctg actgtgaaaa taagggcagg ctacactgga gagcagggat 60
agggacaccc ggggggcaga gatgtgggtc accttagggg aggacacact caggaggccg 120
gcccatgatg gcacatgaag gctgggagca cgggtgctca ggatcagctc atcagggaac 180
ttgaccaaata ttagagcaag gccctttgat agtgtataga gatgtttgtt ctaagcagca 240
atagaaagct tctggaatct gttccattaa gaggtgatag aaacaaaata tgagtcgttt 300
tggagttgtt ttcagcagag tcacaatgat agcaccatta tagatatttt acagacataa 360
tcctgatctt ttgggtggat gaccagaatg tctagttggt tctactgagcc ctgggttttga 420
cccaatatgg taattcgtga actcttagga ggccagaaat atcctaatac tgtgcaaggc 480
agggaccctt ggactgtaac tgtcttgtct gcttttggtc gtgaaggaga ctcagaggcc 540
caaacaagaa tttaggaaaa agagcaatag gattgtgttt aaaaaa 586

<210> 95
<211> 491
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (480)
<223> a, c, g or t

<400> 95

39

```

aaataattta acctaggaaa agaaaaagaa aattgaaaat tggagctaaa ataatttgat 60
ttttccctca acagggttat tggctgtctt ttaagtgact aaaagagcgt atctttatgt 120
gaatttttag catggtcata tgattaatac aaggataaag caaccaaag ctctcagtat 180
ttattcccgt gctatttgtc tgtttttttag ttcattggagt attgtattgt acttggtaat 240
ttgatgcttt tgagatgtcc ttttagacaga tttttaacta caggacttcc tctgtagaat 300
cgacaatgtg tttcactctc tgtggcattg acaatgtttt tgaatgccta attgttcagt 360
agaactccgt gggtattatt acaactttgt acattattat aaatatttta tattagttgt 420
atattccact gcagatagca accagaaaac taaaatacag aaatattaca tattagaggn 480
gattataatg g                                     491

```

<210> 96

<211> 634

<212> DNA

<213> Homo sapiens

<400> 96

```

aaataattta acctaggaaa agaaaaagaa aattgaaaat tggagctaaa ataatttgat 60
ttttccctca acagggttat tggctgtctt ttaagtgact aaaagagcgt atctttatgt 120
gaatttttag catggtcata tgattaatac aaggataaag caaccaaag ctctcagtat 180
ttattcccgt gctatttgtc tgtttttttag ttcattggagt attgtattgt acttggtaat 240
ttgatgcttt tgagatgtcc ttttagacaga tttttaacta caggacttcc tctgtagaat 300
cgacaatgtg tttcactctc tgtggcattg acaatgtttt tgaatgccta attgttcagt 360
agaactccgt gggtattatt acaactttgt acattattat aaatatttta tattagttgt 420
atattccact gcagatagca accagaaaac taaatacaga aatattacat atagagagaa 480
tataatgtac aaaaaaaatc ttgggagatg agtgcttttg gtttaattct atttttactg 540
aaaccagaga ataataggat tcaaatctac ctaatttttc tatttttctg attttccatt 600
ctgtatgctc ttctttgaat ttttccttg gtca                                     634

```

<210> 97

<211> 397

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (326)

<223> a, c, g or t

<220>

<221> unsure

<222> (331)

<223> a, c, g or t

<220>

<221> unsure

<222> (337)

<223> a, c, g or t

<220>

<221> unsure

<222> (371)

<223> a, c, g or t

<400> 97

```

aataattagc caagttgtgg tgctttgagt tttttgagtc tgtgggttaa tatctgtcaa 60
caattttgga aaattatcag ccatattatt tgaagtcttt cttctgtcac atatttcttt 120
tccttataca attagaattg catatatatt agggagtttg atattatccc acagatccgt 180

```

40

gatgatatat ttcattttct tccttttctt tttcctagtg tttcagtttg gacgagtttt 240
atcgacatat ctttaaggtc actaatgatt ttctcagctg tgtcaagtct cctgataagg 300
ccaataaaga gactatatct attatngtgt nttaanttc tagcatttcc attttattct 360
tagagtttaa nctctctaata gaaattaccc atcttat 397

<210> 98

<211> 342

<212> DNA

<213> Homo sapiens

<400> 98

ataaagatgg ggtgagggaa gaaaagatga caaaaggaga ggaccaggca tgagaagagg 60
aagaggagaa tgcggaggag gctgcttgcc tgctgtggga tggatggcag gggcacttcc 120
ccagactcac ttttctcaga tgtaaaactg accagccttg tgccacagat gtgaagatag 180
ccccatagaa cttaaagagc agaccataac ttcccatgaa tgagagctac taacatttac 240
atctgaaaaa caatttggat acttacccaa gtctccaaca aacaaagtca cactgaagct 300
ggagagcaca ctcataacac ccggaaaaac attttttttt aa 342

<210> 99

<211> 873

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (338)..(528)

<223> a, c, g or t

<400> 99

ataaagatgg ggtgagggaa gaaaagatga caaaaggaga ggaccaggca tgagaagagg 60
aagaggagaa tgcggaggag gctgcttgcc tgctgtggga tggatggcag gggcacttcc 120
ccagactcac ttttctcaga tgtaaaactg accagccttg tgccacagat gtgaagatag 180
ccccatagaa cttaaagagc agaccataac ttcccatgaa tgagagctac taacatttac 240
atctgaaaaa caatttggat acttacccaa gtctccaaca aacaaagtca cactgaagct 300
ggagagcaca ctcataacac ccggaaaaac atttttttnnn nnnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 540
aaatcacctg gtgaccattg gacaggcccc agagacaaat cttcttacct gggcaattca 600
gaagggagcc aagaccacct ggtgaccatc aaacaggcca tctggaggca aaactcctta 660
tctggggaat ttagaagtaa tcaaacttcc ctagtatctg aagacggcat ctgatcatga 720
tacaggaact agaaagaaat catttaggca gttagttagg gtgaggggaag agagaggccc 780
tctcatattg tttatttagg ccattagtga gggtagggga agagagagac cctctcatat 840
tgtttcatat tgtttttatac tcagtacctg att 873

<210> 100

<211> 297

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (48)

<223> a, c, g or t

<400> 100

41

ggaaaaggcc cccttaacct tctcctcag gccactcag caaatgtngc cactttgtgg 60
 ccactttgtg taaggcattc cagagatctg gtgaggcacc tatctacaaa tatttatata 120
 cacacattca tatatggttt cagtcacaaa atggggtcac tctctcccct gacctatcat 180
 ttagggcatt ggaacatggc tgcattgtggc tctgtttgtg aggggtccagg ggatggacag 240
 ggaggctctg cattattttg cttttaccaa cattgcagca tgaacgtttt ttttaact 297

<210> 101
 <211> 258
 <212> DNA
 <213> Homo sapiens

<400> 101
 aatataaata cgcctttaat agtaaacacct aattaccta caccatcaaa aatgggggtgc 60
 tccatgaaga agcacataat tcaaattatt gaagtttatc ccttctaata accacataga 120
 tttctcttgc ccattataaa aattagataa tcagtatttc taggatagtt gttttcttcc 180
 aaccaattaa ggcataatct atgtagcaga acattcagag gatgatgcct ggtcaacatt 240
 tgaataaaca atcactgt 258

<210> 102
 <211> 712
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (603)
 <223> a, c, g or t

<400> 102
 aatataaata cgcctttaat agtaaacacct aattaccta caccatcaaa aatgggggtgc 60
 tccatgaaga agcacataat tcaaattatt gaagtttatc ccttctaata accacataga 120
 tttctcttgc ccattataaa aattagataa tcagtatttc taggatagtt gttttcttcc 180
 aaccaattaa ggcataatct atgtagcaga acattcagag gatgatgcct ggtcaacatt 240
 tgaataaaca atcactgtga tgttacctct atttaagatg actccaataa aacttctatg 300
 gtttgcatta ttagttgatc agactttaag cattatcttt tgatagggtc aaggaacctg 360
 tcttaactcc ccattctctga ccaaaatata cttgttttct ataagctata aagccagata 420
 gccaatTTta tgagaattgt cctataacta tatccatgtg agcgatgagt gcctggcatg 480
 aagatgcata aaggaggcag taatatacaa caactgaagc ataacctctg gagccagtct 540
 tcttcagaca aatcccaatt ccattactca ctggccacct aaacaagcta ctttaattcat 600
 ctncctcagt tttcttcaac tgtttaatgg gtatgatcaa caaaccaact tcagtgggtt 660
 atcataaata ttaataaatg agagaatgca tgtgaaacaa agctataagc aa 712

<210> 103
 <211> 173
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (96)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (140)
 <223> a, c, g or t

<400> 103

```

gaatgtggct ggtgagtagg cacttgggtgt ggcagtgtgg ctagtgggta agaacatggc 60
tgggtgattag gcatgtgggtg tggcagtgtg gctggngggg acgagcatgg ctgggtgggta 120
agaacgtggc tgggagtagn agcatggccg gtggttggga atgtggctag tga 173

```

<210> 104

<211> 688

<212> DNA

<213> Homo sapiens

<400> 104

```

tctgaatggt ttggtgaata aatctgttct tcagcaaccc tacctgcttc tccaaactgc 60
ctaaagagat ccagtactga tgacgctggt ctcccatctt tactccctgg aaactaacca 120
cgttgtcttc tttccttcac caccaccag gagctcagag atctaagctg ctttccatct 180
tttctcccag ccccaggaca ctgactctgt acaggatggg gccgtcctct tgccctcttc 240
tcatcctaata ccccttctc cagctgatca acctggggag tactcagtgt tccttagact 300
ccgttatgga taagaagatc aaggatgttc tcaacagtct agagtacagt ccctctccta 360
taagcaagaa gctctcgtgt gctagtgtca aaagccaagg cagaccgtcc tcctgccctg 420
ctgggatggc tgtcactggc tgtgcttggt gctatggctg tggttcgtgg gatgttcagc 480
tggaaccac ctgccactgc cagtgcagtg tgggtggactg gaccactgcc cgctgctgcc 540
acctgacctg acaggaggga ggctgagaac tcagttttgt gaccatgaca gtaatgaaac 600
cagggtccca accaagaaat ctaactcaaa cgtcccactt catttggtcc attcctgatt 660
cttgggtaat aaagacaaac tttgtaaa 688

```

<210> 105

<211> 977

<212> DNA

<213> Homo sapiens

<400> 105

```

ggcttggaga gggtcacaga ggctagtagc tgtgtggact tgcaggcagc cccaaatgct 60
cacctatgtg cagagtcagc atgtcctgcc tcccctggta atgtggtcgc ctgcatctct 120
gtggccagcg ctctcgttca tcattcagtc tgatggcttg agtgcctcta tgtttgctac 180
atgctgagac cgtattctag tgccgtatc tggaggtagt ggggtgtacct acagatttaa 240
gaatgcaaat ctggaggtac acccagtga ttcaaagtag tctcatagaa caaagagact 300
tatatagtga cctttgctgc atccactagt atacaccatc tgaggctctc tgaactgaaa 360
atgaatgtgg aagcaaggga acagtgtgat gttcagctct cagatctcac atggcatctg 420
atttggcttg aggtgcctcc cctcctctct gtcccctggc tgtgggctca tggattggca 480
gagcccagtt atggcttcg ttttacttgc tataatatcc agaggcaatg tactagtcta 540
cctagaaaat tgtgtcacg gcaccccttt gtcacattaa taagcattat ggacactacg 600
acattttatt aagtattttg ttctggatc tacttgatta tagtaaatta tcaaaatcct 660
tatttagctc atggactctc attaaagcat gttctggaaa ccttggccat aggttaggag 720
cctgtaaagt ttgattcatt gcaagatata agtgattagc agttggtagt agtgacattg 780
atgggcccga ttaaaaggtc tattggatgt ggtgggtggc tagcgatagg ttggagtgg 840
aggtcagcat ggatgtctct gatctagaac caagcttacc tttgcataac ctatagtac 900
actctcttca tctccccacg ccttagccat gtctccctga gggttcatact gtttgaatt 960
tcacaggctc atttacc 977

```

<210> 106

<211> 500

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (357)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (367)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (391)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (410)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (430)
 <223> a, c, g or t

<400> 106
 cagagcaggc attgacctag atgtcttccc ctgccttcat tgggaggggt ctgagccacg 60
 ggttccacct ctgccaaagg cacacctagg agactcctca tgtccagctg agaagagggg 120
 gacacctcct gtctgagact gcagctcaca ctgctgcatg cttcctggac accatctctc 180
 tgaccttggt cgcctctgcc tagcctgcag ctacgttctc tgacctccag ctcttctctc 240
 ttctccccctc ggtaataacca aagtctcaag aacacagccc tcacttctag acagaaaggc 300
 ctcaccagga cccacctgtg tggcccagggt gtgacctcat gtacaaacac atctccnaaa 360
 atcacctctc cgtcatcatg gaccctagta ntatccatga gttaacnctn atttctgtgt 420
 taatcggggn tgcagcacat tttggtgcag attcattgtg gctttggggg gccatttggg 480
 actctcccc atgcacaatg 500

<210> 107
 <211> 476
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (466)
 <223> a, c, g or t

<400> 107
 gccatctttc cactcattcc ttctcaaaag gaatgtagta ccatatagta gttaagaata 60
 tagacactgg agccgatctt cttgagttcc aatagtggtt cttctacttt ttaaatctca 120
 ttttccttca tctttaaatt gaagatagta acaatctcat ggggttggtg taactaaggg 180
 ggtaatgcat gtaaagtgtc tagaaaatgc ctggacatag gaagctctaa gtttgctgct 240
 actactgtta ttatgggttac tattattaat cattgcaagg aaaatgtatc aacagatgaa 300
 tttggttcaa tactgccttc tagttttgtg accttagaat ttataggaac aaaaaagatt 360
 tgaagggagg ttgggctgga tcatagagag ctttgattcc atgttttagg atgtatacac 420
 agtgagaagt ccttcagggt ttggtcctgg gaagagttgt gaatcngaaa gttaac 476

<210> 108
 <211> 834
 <212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (824)

<223> a, c, g or t

<400> 108

```

ataagtatgc atgcttcata tacttcattt attcttttctt ccttgaagcc tctcctcttt 60
attaggcact attcatttgt ctacttggta cctgtatttt tttaatgtca ctattttgac 120
agtaccaata aaggtaaagc cactcaatta cgcagggctc tctctttatg ctttgggtag 180
gtgcacctgt gcaactgagg ggacggtcag tgttatcaag gttacctgtt attacaagta 240
gaagaaccca caaagatcag gagagagctc attttcctcc attagtagga ggtaggacta 300
tacattcaca aacacgaacc ttaaaatagc tcacaaaata gtgtcatata tgtaccacag 360
catctttcca ctcatctctt ctcaaaagga atgtagtacc atatagtagt taagaatata 420
gacactggag ccgatcttct tgagttccaa tagtggctct tctacttttt aaatctcatt 480
ttccttcac c tttaaattga agatagtaac aatctcatgg ggttgatgata actaaggggg 540
taatgcatgt aaagtgccta gaaaatgcct ggacatagga agctctaagt ttgctgctac 600
tactgttatt atggttacta ttattaatca ttgcaaggaa aatgtatcaa cagatgaatt 660
tgggttcaata ctgccttcta gttttgtgac cttagaattt ataggaacaa aaaagatttg 720
aagggagggtt gggctggatc atagagagcc ttgattccat gtttttaggat gtatacacag 780
tgagaagtcc ttcagggtttt ggtcctggga agagttgtga atcngaaagt taac 834

```

<210> 109

<211> 498

<212> DNA

<213> Homo sapiens

<400> 109

```

tttaaattgg gagttaagga tgagcacttt tactgtatta aaaaatactc accagttaaa 60
aaaaaatact cttttccctt tctcggaca cctaaatcta agagaacaac tcctatataa 120
aaatgatata aaaatcatac attttgggaag tatgtttcta actgttctga gaggtgcat 180
ggtaaagctg aagtgaagaa tgtattttta atctgtatat atgagcaagt atatattgat 240
gattgaagct aggtgctgcc taaatacatg gccagactt tgaggaatta tagtgtaatg 300
gttggaata caggtttgga gtcacaccgt agagctgaaa gcttggcttt tatttagctg 360
tgggtccttg ggcaggatac gtaatctgtc tgtgcctgaa ataccacca caccatcct 420
gtaatggggg gataataagc ctgcctatct catggggcta ttaagaattt tcagttaact 480
tttacttatg aagtgcata 498

```

<210> 110

<211> 259

<212> DNA

<213> Homo sapiens

<400> 110

```

tttaatgtgg tttagtttta gtcacttaga tttgcttttt atggagtgac tggagtttgg 60
ggagggggagc agggagggtt ttcttttttt ctttataaca ctggctaaat attttaatta 120
ctgctataga aggaagaagc taaaagtatt gcattcaca atattgcata gattatacaa 180
acacagaaat atatgcatat gcatgtttta aatatatgcc acatatcaac accatgtatc 240
caacttgaat aaggtcatt 259

```

<210> 111

<211> 414

<212> DNA

<213> Homo sapiens

This page is not part of
the pamphlet!

WO 02-48370
3/3

Date: 20 jun 2002

Destination: Agent

Address:

45

<400> 111
 atgaaagga tgaggggaac tcaaagttac aatgtcctac ttggagcagt aagttcagta 60
 gacatatcac ttgcctcatt aacatcaagc atcccaaac ccagtctggg tcagttttgc 120
 ccagagtggg gttttagtaa cacgggttct cctgggatcc tatacctagc ccagaatcag 180
 ttgcaaaagc caggccatag caaattgtcc tgccagccag atagcagaga atctgacggc 240
 agcaggcaga aggagccgct ccattgcagt aagccaagat cgcgccactt gcctcattac 300
 atcaagcatc ccaaaaccca gtctgggtca gttttgcccc gagtgagggt tgtagaacac 360
 gggttctcct gggatctata cctagcccag aatcagttcc aaaagtccaa aaga 414

<210> 112
 <211> 589
 <212> DNA
 <213> Homo sapiens

<400> 112
 ctgggcaaca ttggggagac tctgtctcta aagaaaaaaa ggagagctgg tggtgaaagt 60
 gtgaaggacc caggaagtac agacactggg ggtcaaagaa caagggtagg agtgtcatca 120
 aatgatatgt ttggcagcat gggagctgtg ggtagagagt gagataccta aatttatgat 180
 ttctgggtgg cagtaacttc taggggtgtg ctgtgggagt gggcctctga atggggtgga 240
 ggagaaaatc attaaagatt agaaaatctt gggattttaga ggatagggtt tgggatgggt 300
 gatacacgtt agtgttgcac ttgcccaggg taacgccaaag agttggcaga gaaaataata 360
 ctgacctaga ctttaataaa ggatttggga atgacagaga agcaacagta aaaataaggg 420
 ataattagat gtttgggtgt ttgcctggc tgtgtctgtc ctgtgtctgg ccaattatta 480
 caatgtattt acactgtaaa tacatgtaac tcatataata gttttataag tagcaaaatg 540
 tagtttaata aaaaaccatc ttagtcttct tacagaatat ttagttacc 589

<210> 113
 <211> 471
 <212> DNA
 <213> Homo sapiens

<400> 113
 cccaggctgg gggtcagggt aggagggagc tgggatccag caagcctagt gaaacccagg 60
 ggacagtggg ctcggtcaca tccaggatgg tgatcaacag ctgcatcatc ccgttcctt 120
 ctcaagcgac aattccagag ccttggccac acgggtgctt tatctttcgt attcagaccc 180
 cctgggggttc cagcccctta ctgccttcac ttctctctca ccccttgact catctttcct 240
 gctacttgtc acttgagata cctaagatga tgtgtgttat ggagagggtta gaccaccagc 300
 ttcaagaacca ccctgtgact ttggcctagt cacctgacat ttctagactt tgggtgtctc 360
 attcataaag gcagtgtgga ctgcttgctg atgttatcgt gaacctgaat tccttcttag 420
 agtttctaag tgctttctgg ggattaacct tttaaatcct tgcagtagcc c 471

<210> 114
 <211> 1032
 <212> DNA
 <213> Homo sapiens

<400> 114
 aatgagggag ctcttgagct cccttgatga gcaccacaca gggccctctg ggaagcagta 60
 agaacccatc ccagggtcca ataagaacct aaccagcctt gggatggccc ttccctttct 120
 gccaaagtcc ttcccatgac aaacctcagg cccttatctt ggtatctgtc accaccacc 180
 acccccccga cacacacaca gtcatgcaag ttgtaagaca gtgacagaag atttgaagaa 240
 gaccaccaga gcaggggata gcagaacatg cagacttagg gggaaagccag gcgttcatac 300
 caaagaatta gacctgttgg gtaccacaggc tgggggtcag gtgagggagg agctggggtc 360
 cagcaagcct agtgaaaccc aggggacagt ggactcggc acatccagga tggatgacaa 420
 cagctgcatc atcccgttc ctctcaagc gacaattcca gacccttggc cacacgggtc 480
 ttgtatcttt cgtattcaga cccctgggg ttccagcccc ttactgcctt cactttcctc 540

46

```

tcacccttg actcatctt cctgctact gtcacttgag atacctaaga tgatgtgtgt 600
tatggagagg ttagagcacc agcttcagaa ccaccctgtg actttggcct agtcacctga 660
catttctaga ctttgggtgc ttcatccta aaggcagtg ggactgcttg ctgatgttat 720
cgtgaacctg aattccttct tagagtttct aagtgtttc tggggattaa ctttttaaat 780
ccttgcaagta gccaataaag gtaggtattg ttgttatccc cattttacag gtaaggaaac 840
tgaggcacag agagtaattt gcacaaggct tatggctttt tagtggagga gccaaagagtc 900
aaattaagag tggttgagtc aggcattgtg gccctgcct atagtcccag ctacttgaaa 960
gagtgaagtg ggaggatcgc ttgagcccag gagttcaatg ctacagagca agacctcaac 1020
tctttaaaaa aa 1032

```

<210> 115
 <211> 440
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (428)
 <223> a, c, g or t

```

<400> 115
ggactacatc catgttccac cacaccaggc tccaattaca ttttgacttc tccacttggg 60
tgtttaaaat gcttctcaaa tttaacatat cctaaagata attttgtgtc tccccacaaa 120
acttgcctct tttgcattca ttgctgtctt agttaatggc accaccatcc atactgttac 180
tttagccaga aacctttgaa acatcccaat tggcttttct gattttctct gtttcacaa 240
ttattctcca cagacaggat actccaaaca gtacccaaag ccattgtctc ttatactttt 300
caatctataa aatatacata cataagagta tataaaatat attataaagt aaatatccat 360
gtatccaaac acacagggtt agaactggga acacaatatg caaaagaata atattgggac 420
ccccctancc tcatgtcata 440

```

<210> 116
 <211> 249
 <212> DNA
 <213> Homo sapiens

```

<400> 116
aaaaaaagtt ctgacaattt gtttgctttt acattttcaa atttgtgaaa ttagagagata 60
attttgtttt caaatctttg taattccctg aagcaaatac tttcaagcca gttgcaaaat 120
gctgctttag aaataattca tataaacatg cttctctatt taatcacaag gggagatgtg 180
gagaatggat gttttatttt ttcagtagtt tttgctctat aaaaatatta aattgctatt 240
atgattact 249

```

<210> 117
 <211> 1017
 <212> DNA
 <213> Homo sapiens

```

<400> 117
gccctttttt ggtgtgcccc ctgaatgagc actccaggct gtggagtctg ggacatgcct 60
tgggtttgtg ggaccatgct gcctgcctgt cgagaccaag catcgatact gtgtgtctac 120
ctgatgaaag tgtccagtat gtgtctgcat gacttgggga cactaagaaa accaaagggg 180
ttagcaacaa agagagcttg tcacctttgt gcggaaccag ctggcatctc acagggacaa 240
cctacaacct gagctgctgc gtccctacta aatctgggcc cctagggacc ccgttttact 300
cctgctctcc tggagcttat tacgggcctg gctaccaaag ggaaagaggg gaaaatagac 360
caggagcctt atgctagaac cattttattt gtttcacgtg atgcagacag agataaaact 420
gcaaatttaa tgaaacttta acaatcagta caatgtttct ccttaagaac tttgtaata 480

```

47

gcatttatct ttcaagagtt ctttctctct ttttgtgatt attttataaa cttaaaggaa 540
aaagagaaaa agtcagtggg tccagcattt gcttttagtct gtgacttaaa tggattataa 600
ctcttgaccg ctgacattta ccaagataaa tcagtgggta tagatgtgga gcttgatgtc 660
tcttcggctc tgggaccaat ccccttggaac aaaagtttct ctgtgttctt agtattctga 720
actggctaca gcaactttta ggaaaataaa gggtacaaaa aaagttctga caatttgttt 780
gcttttacat tttcaaattt gtgaaatgta gagataattt tgttttcaaa tctttgtaat 840
tccctgaagc aaatactttc aagccagttg caaaatgctg ctttagaaat aattcatata 900
aacatgcttc tctatttaat cacaagggga gatgtggaga atggatgttt tattttttca 960
gtagtttttg ctctataaaa atattaaatt gctattatga ttactaaaga taaaaaa 1017

<210> 118

<211> 332

<212> DNA

<213> Homo sapiens

<400> 118

ctgectccac gtggattacc acatttctca cctcatccta caaggcagtt cctgtttcta 60
ttcccccttc acacaaaata acttcgtatg ttgttagtaa gcaggagAAC cagcctttga 120
actcaggact gtttaaagac caaggctctg gccactgaaa taaaacatct gcaactggca 180
gattaatgaa aggtctctaga aggaaacaaa aaaccaaga gactgctggc agtगतtagct 240
gagttttagg gggaaaagtt gtttttagttt tccctgtata ctttcttggt tagtttttaa 300
aatctacagt atttacactt tcaaaacaaa at 332

<210> 119

<211> 344

<212> DNA

<213> Homo sapiens

<400> 119

gcgcagggga aattataggt ggctgtgggt gtaattacaa agttctgtca cgtcttcatt 60
gttaggagga aaagaattca ataatectat cagttctgct gtaaaacaaa tgagctatga 120
aattctgggt aacactgatt ttatgtctcc attcttgagg aactgttag tttgttttca 180
tctgtatgcc ttgattagag caaataacct taaatacct taaggaaact tagatataca 240
tcatttccag tttttatcaa atgtgaattt tttttgtcat actgcccacc taacatggga 300
tgttttctca gaattattgt cacttatgtg tttgagtttc ttaa 344

<210> 120

<211> 718

<212> DNA

<213> Homo sapiens

<400> 120

aaaaaatcat aatagtttat gatcttgaag ggtttaaaag tatttgatga agatgtcttt 60
tgaatttatt ttaggtctt cttgtgtatt taaaagctaa gttatcttgt aatcattttt 120
ttctatacct ttgtcagtaa cctcttagtg atgaaataaa aaagattagg taatcatcca 180
gcaatgggga agaagttaag gaacaaagag ctgagattaa actagttttt agaactcaag 240
catttctgca tgaatttgaa tcatggaaaa caaatgttag cactccaaca tttgatgcaa 300
aactaaaagt ggaatactgc tttgatattt gaatgaattg aaaaataatt aacatccttg 360
gaactgtatg taaagaagga cttcacagt attatagata cccccaacct cagccctttt 420
cccatgtatc tctttgatca catcctacc tcatagatca cccatgtgct gaagactttc 480
agttctgtat cttcattcta gatctcctga actcaagatc agaatactt tctgacttct 540
gactgtgtat ttctggatgt tatacaagaa cctcagctca aactcagtat tccctaaacc 600
attgtttttg aaactttatg ttggatgtga aatctgtatt gtagaataac attaaaaaa 660
gaaagaatag tatgcaaaat atcagagtgc attgtatgta gcaagagtag gtattttt 718

<210> 121
<211> 2617
<212> DNA
<213> Homo sapiens

<400> 121
atgtggaatc aacctacctg tccaggaaca gatgaagaga taagaaaatg cagtgtgtat 60
acacagtggg atgctcttca ccataaaaaa ttcacggaat catgtcattg cagcaacatg 120
gtggacaatg taagaaaagc tccccggaga agctgtacag aagctgcctc ctcagcagtc 180
agggccaggt accggagctg tttttacca aggacagggc cggccccaag tcatcccaga 240
gctgccatgg caccacctca gtccgggcct gaggaatcct acacaagcta cttatatcag 300
tgatcactag gataatccat agaacttttg ggaaagaagt ttaagacctt tctcccacca 360
tttcagcagg ataaattcca actggattag aaaatgaaat gttaataatg caaataagta 420
catatttata tctgtatata aaatacagtt gatatttgcc tgggtgttag gtgtctaaag 480
gactttttaa gcataaaggc aaaaaaaagt cataaaaatg ctatagcagt ttgagactct 540
atgcaggaaa gggcatcatc acgtgcatgg atgaatctgt atctaatttt aaacaatttc 600
caatgggtgc tgtttccttt tctttgaaaa tctctggaga aatagttcct cttgctgtgt 660
ctttcttttag gcaagaattt ttactaattg atgtgtagtc tgaatcctgg ctaagtataa 720
accttttatt ttttatacct gttcttagtg aaaatgaaac tgtgactttt tttttaatc 780
cttttggttg tcaaaaacta caattaactc ttctgagttt cttctctggc tgaacaaaca 840
atgggtcccat tggcctttca gggaactcca ggccgtctca aaaaccttca tgtttcattt 900
cttttcagag ctcccaaaaa gaatagcttg ctcttgacgt tgtacatgtt agtggaatga 960
tcaggactac tttgcaaaga tgaaaaattt gtgtttctag tgatttgaaa atagaaatct 1020
gatgtaacta ttagatattg ggaagaagg tgacgaagg aggtatcacc gaaagcactt 1080
aacaattctg aataattctg tacttgattg catttatgtg tatcatagga acagttgggt 1140
ttccttgagt gttaaattat ttattcactt attccacttc aagccagcta aatgattgtt 1200
tccttgatgg caaaagtctc agattgattg cacagttaat ttgggtggat tggttatgtc 1260
ctttttatta tttattctta tttcaccaat gaaaatatca ctaagttcct tggtttggtt 1320
acctgattgt acctactttg acaaatcact gcctttcttg acccagtttt ctcattaagt 1380
ggcagtgata acctgtcata cttacagata taaaaacatg aaagttaaag tattgggtaa 1440
tactttcctc ctatcttttt tttattttga aaaagataaa aaattggcat aatgtattag 1500
ttaagatgga ataatacatat gttgatatcc agccatttct tctctcaaat gataggaaga 1560
tttttatgtg aaactacttg tgagagatct taacaatttg tagtttagaga aagcactatt 1620
atatcatttg gaaatgcaag aaacaagtta cctttggggc aacagaggcc cttgtcattt 1680
tctcaaaaga aggaagcatc agcattttga tgatgatgtt gagattgtag aaatgatgaa 1740
gggtgaaaag ttattctagc ttatgttttag caaaatgaaa tgaacccaaa taataaaaaca 1800
gttacaacat tgaatctctt tgggagaaaa aaaaaagata gaatgctaag gtccttcaga 1860
acttcttaaa ccagaacctt aaaaaaaga gaagctttta aaaaatcata atagtttatg 1920
atcttgagg gtttaaaagt atttgatgaa gatgtctttt gaattttatt gtaggcttc 1980
ttgtgtattt aaaagctaag ttatcttgta atcatttttt tctatacctt tgtcagtaac 2040
ctcttagtga tgaaataaaa aagattaggt aatcatccag caatggggaa gaagttaagg 2100
aacaagagc tcagattaaa ctagttttta gaatctaagc atttctgcat gaatttgaat 2160
catggaaaac aaaatgtagc actccaacat ttgatgcaa actaaaagtg gaatactgct 2220
ttgatatttg aatgaattga aaaataatta acatccttgg aactgtatgt aaagaaggac 2280
ttcacaagta ttatagatc ccccaacctc agcccttttc ccatgtatct ctttgatcac 2340
atccctacct catagatcac ccatgtgctg aagactttca gttctgtatc ttcatctag 2400
atctcctgaa ctcaagatca gaatatcttt ctgacttctg actgtgtatt tctggatgtt 2460
atacaagaac ctcagctcaa actcagtatt ccctaaacca ttgtttttga aactttatgt 2520
tggatgtgaa atctgtattg tagaataaca ttaaaaaaag aaagaatagt atgcaaaaata 2580
tcagagtgc a ttgtatgtag caagagtagg tatttttc 2617

<210> 122
<211> 373
<212> DNA
<213> Homo sapiens

<220>
<221> unsure

<222> (74) ..(294)

<223> a, c, g or t

<400> 122

```
gtattataat aatggcctta atgaataaca ttctctatat tcacacttat ttgcaatata 60
atactgccat tctnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnncactaa 300
tctcaaagag ggcattgatct tcaagaatta ataaccctct caagtctcta caatctaattg 360
caattacctt ggg                                     373
```

<210> 123

<211> 308

<212> DNA

<213> Homo sapiens

<400> 123

```
gctgaaagcc cagagcagag ctgttctcat ggggaaggac cctgtcttcc ccatcatcct 60
aggcgttcat tgaggatgag gactgtcttc ctccatcaga ccgagagtcc ccaaggggcaa 120
gggctgtctc tccctgggca gacagggagc tccccgaggg cagagggtcct gtctcctcca 180
tcagactggt agccccaca accacaaagc tatgtctact ttcacagaa ggagctccct 240
aagtggggaa gggttctccc tattttccc ttccaggtgg gaaattcctg gccagggtcc 300
cctgtctc                                     308
```

<210> 124

<211> 774

<212> DNA

<213> Homo sapiens

<400> 124

```
gccaacacca aagggggcac gggagaagga caggaggggt ggtttccctc agcaagctct 60
cagtcccact gacactggcc caagagggtc gactgtactg ggcactcacg caggagagatt 120
gttcccgaag gccctcgga aagttggtga atgcaaacag caggcagcca gagagcctgc 180
tgcagaggag accagagacg atgcccagc agggcacaga agtgtgcaaa agactcagca 240
gtgggaagga gcctgggtccg tgagtgtgag gagataaccg gggccctagg cccttcctgc 300
cccaactttc caccacctgg cccagcccct tgcagcgtg aggccttagca tctctctgct 360
gggtttgtga gagcccagac tgccccagtg aggttacagg agtactctcc ccaggcagga 420
agggtgggag gcctccctcc aggtacccaa gaggaatgt tagcagctga aagccccaga 480
gcagagctgt tctcatggg aaggaccctg tcttcccct catcctaggc gttcattgag 540
gatgaggact gtcttctcc atcagaccga gacttcccaa gggcaagggc tgtctctccc 600
tggtcagaca gggagctccc cgagggcaga ggtcctgtct cctccatcag actggtagcc 660
cccacaacca caaagctatg tctactttca tcagaaggag ctccctaagt ggggaagggt 720
tctccctatt ttccccttcc aggtgggaaa ttctggcca ggtcccctg tctc 774
```

<210> 125

<211> 271

<212> DNA

<213> Homo sapiens

<400> 125

```
aagtcgtacg catggttaaa aaaaaaagaa aagaaaatcc aaaatagtag tgaaggtagt 60
cagtacacag gaagcctccg cccacctcca cctcccagct tcccccttg gaggtatctg 120
ctgtagtggg ctctcaaga tacttctagc catgctctgt ttgtgcatgc ttatccctgc 180
acagacagca gaagctgtct tggccaacaa gaccaggaag cattggtatt tgcagggtaa 240
ttgaaaaatt catttaaggt ggagaacccat a                                     271
```

<210> 126
 <211> 1950
 <212> DNA
 <213> Homo sapiens

<400> 126
 atgatgccac aggatgagcg cacttcaaag ctggaaggaa gcctggtgag ggagcagggc 60
 agaatcttct cctggactgt gagggtagat acggtggatg tgtatggctt cattgaagat 120
 gccagtcctt gcattggcat ctgcagattt gaagaagtag gcccctcttc tagtcttcat 180
 ggactggatt tggcaagaaa agtccttcat cagtcagcca ttcagaaact ctgggaagcc 240
 tatctggtaa cgtccatggg caggcaaaat ttgccattca gctacaagaa gtgcagttag 300
 cagacagcct tcaacttcag catcttcaga gtctgccttg actttcaagc tgaggccatg 360
 gacttctcag gagctcctag ccaatggctg agaacaacgt gtctaacaca tgttctcttt 420
 ctctttgatg gccaaggcat ggctggccaa tgggatgctt ctctctccaa aggagcaggg 480
 agagctggag ataccctcct tgcaaacagc agcttgagga tccagcgcct ggtgcacagc 540
 ccacagcgac cccaagaagc tgctccaacc cctgggacta tggagctcta cagctgtaga 600
 gaccaccagg aagtggactg caggcccctg gcctctccat tcagattctg caaagagatc 660
 ctgatgggtt gggccaatgg gtcaggcatc cagtcagctc tggctaaggg agctgcctgg 720
 tgccaggagc agcgtaacac ggaccacagc tgtccccaga agggggcagg cgttctgaga 780
 gccacaaagt cctggctgcc agtgctccct ggtctgatcc taaaccgcgc ctccctgggt 840
 gacagcttcg ccgtgagcgc tgcctgggct cggaagggca tgcaggagtg gatcgggaga 900
 cagcgtgcc cgggcgggtg ctccgggacc cgacagctgc ggttggcggg caccataggc 960
 cgaagcacc gggaaagaga ccctgagacg ctgctgaatg agcaaagcaa ctgcaaaaaca 1020
 ttcataggcc atggctcctgt ttcttacagt gtgaaaaagt ctattcaggc ctgtgtcact 1080
 gtgtatctgc agatgggttg atcagagcac ctctctgtga tgtcacaaat cggggccttt 1140
 ctagecctct taaccttgga ggttctgctc agcagctgct actggcgtct cgtcctcttg 1200
 gctctgggtc tggggcactg gaaggtaaac tccctgctga gttggaggca gcagcattga 1260
 gtgggtggct gttttccagc caggatttac ccagggcttt atggcttgca aagccttctc 1320
 cacagggtct tgcaggcat ttaattattca caaaaatgtg gccaggatca aaattattat 1380
 tatggggaaa ctgaggccag actgtaaagt ccacaggcca ggttctttgt ggctcactct 1440
 tgtatccctg ggcttttgc actgattggc acatggcaga tcctcaagaa cattttccag 1500
 gtggatgagg ttcagagggg ccatgcagct tggccagagg gcacacagcc agagaggcag 1560
 ggattctgtt ctgttctgtc caagtcacca cctcttttat ggagccaggc tgttctgtgt 1620
 ctttgaagag agcctctgcc ctccagaaag ggtcctcacc ttttctctt ctgtaaat 1680
 agtcgtacgc atggttaaaa aaaaaagaaa agaaaatcca aaatagtact gaaggtagc 1740
 agtacacagg aagcctccgc ccacctccac ctcccagctt ccccttttg aggtatctgc 1800
 tgtagtgggc tcctcaagat acttctagcc atgctctgtt tgtgcatgct tatccctgca 1860
 cagacagcag aagctgtctt ggccaacaag accaggaagc attggtattt gcagggtta 1920
 tgaaaaattc atttaagggtg gagaaccata 1950

<210> 127
 <211> 209
 <212> DNA
 <213> Homo sapiens

<400> 127
 gttgggtgtg gtgggtgtt ttgttgttaa tgttgtttt gccagtctgt gttgataaga 60
 tttattattg agaatagtgc ttgttctctg agtactcctg acttagaaaa ggagcatagc 120
 cctactaaag gggacttcaa agtagaaatc gtcaataacc ttttacttgc tacagttagt 180
 ggctcaaca tgatgtttt aaagatctt 209

<210> 128
 <211> 496
 <212> DNA
 <213> Homo sapiens

<400> 128
gcctccggtg gatggaatga agcaaggatg ggggctgcct gcagagctgt gtcactcact 60
tgtattcagc ttctctgcct ctggctctct gtcttttacc nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
ctaagaaaac aatgatcacc atacatgctc tgcttccaaa ctatactttc acatccaaag 480
taaccccaga ttcata 496

<210> 129
<211> 252
<212> DNA
<213> Homo sapiens

<400> 129
catttctaac atttattgtc ctccagtaca aagaagtaac ccattgtcat gtctactcta 60
tgataggcta gaactatagg gttgctctat attgatcagg tttttaaaaga taaaaatgaa 120
aaaaaaatcc tatccagaca aaataaatca gtgttttata tttttggagc atcagaactt 180
actttaagac ctactggta attcttttagc ctctcacatg tgataaagac attgtgctta 240
cattttttta aa 252

<210> 130
<211> 149
<212> DNA
<213> Homo sapiens

<400> 130
atcagaatcc tgggaagggt ttgttaaaac actactaggc aggggtgaggt aacctaagag 60
cttttgagg cccagggtgag agggatcact tgcggccagc agagttcaag agcagcccag 120
gcaacacagg gagacctctt ctctacaaa 149

<210> 131
<211> 390
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (217)..(273)
<223> a, c, g or t

<400> 131
agcaagtacg cagcattggg aatgaaccaa actcgtagga ggcacagccc actcagtgtg 60
cgggcccggg cgagctgcag gcctgaaacc caccaccct cttagatgtg tctgtgggcc 120
atagaaatta ctagggttgt cttgggtgtg gcctcaacct gttcaacaac aggtgtgctg 180
tttccattct ggaaaccagt cctctgtctt ccagaannnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
gcttttgagg gccaggtga gagggatcac ttgaggccag cagagttcaa gagcagccca 360
ggcaacacag ggagacctct tctctacaaa 390

<210> 132
<211> 1079

<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (874)
<223> a, c, g or t

<220>
<221> unsure
<222> (879)
<223> a, c, g or t

<220>
<221> unsure
<222> (885)
<223> a, c, g or t

<220>
<221> unsure
<222> (887)
<223> a, c, g or t

<220>
<221> unsure
<222> (890)
<223> a, c, g or t

<220>
<221> unsure
<222> (894)
<223> a, c, g or t

<220>
<221> unsure
<222> (896)
<223> a, c, g or t

<220>
<221> unsure
<222> (899)
<223> a, c, g or t

<220>
<221> unsure
<222> (921)
<223> a, c, g or t

<220>
<221> unsure
<222> (924)
<223> a, c, g or t

<220>
<221> unsure
<222> (926)
<223> a, c, g or t

<220>

<221> unsure
<222> (931)
<223> a, c, g or t

<220>
<221> unsure
<222> (933)
<223> a, c, g or t

<220>
<221> unsure
<222> (944)
<223> a, c, g or t

<220>
<221> unsure
<222> (950)
<223> a, c, g or t

<220>
<221> unsure
<222> (975)
<223> a, c, g or t

<220>
<221> unsure
<222> (977)
<223> a, c, g or t

<220>
<221> unsure
<222> (988)
<223> a, c, g or t

<220>
<221> unsure
<222> (993)
<223> a, c, g or t

<220>
<221> unsure
<222> (995)
<223> a, c, g or t

<220>
<221> unsure
<222> (1007)
<223> a, c, g or t

<220>
<221> unsure
<222> (1013)
<223> a, c, g or t

<220>
<221> unsure
<222> (1030)
<223> a, c; g or t

<220>
<221> unsure
<222> (1037)
<223> a, c, g or t

<220>
<221> unsure
<222> (1050)
<223> a, c, g or t

<220>
<221> unsure
<222> (1061)
<223> a, c, g or t

<400> 132
gggatgaaaa cttcctttaa aagaatcctg ttgtatttta atattgttcc ggggttcttt 60
gcatatgtat atgctctata tgaacaatac tgaaatgaac atccatatct atgacctctc 120
tctgcactcc aggctcagat atgcaactcc ctatttgaca ggtctgcttg aaaacttgct 180
gggcatccca gaggtaacat ggatctaata gaaggtttga ttttgcctc caagccagtt 240
cttcccttga ctttctacat ttcaccaa atgaccccaa ccaactcatt attctagccc 300
aagatctagg agttattctt aggttttctt ttacccctc cacatggatc catcagcagg 360
tcttgttctt ttttcttccc aaatatatct caagtccatg ctcttctgtc tgtccctact 420
gccactatcc aagctctgag gccatccatt acatggacaa ctataaacta catgtcctaa 480
tgacatatata gcagtagagt tgctagggtca aaagatttgt gtgttttatt ttgatagact 540
ttgctacatt attctcaaag aggttttctc agtggtatct gcttattata tgagaatttc 600
tgtttctgta ctctgtcacc accactgaat atcagggtca ctcttagccc atagcctcgt 660
gagaattaga agtcacttcc tctgggtgag gcagctagct ccacagcaca gacttaacaa 720
gtggaacttt agcatgtatt taattcccac tcattctctt acctatgtgt ccttctgcag 780
tcaacactct acacaactgt acatgaccac aatgctgtgc ataaataatt ttttagactc 840
tttgtaaata tatatgtaaa aaatggcatc ttantttgna taagnanggn ggangncant 900
taaaattcct tttccttgga ntgncaaat nanagacttt cctnattttn agggttccta 960
acaaattgga aaatncnggg gttaaccnaa ggnatcat atatttnacc atnaaaaatt 1020
ttttcctggn accttangtt tgtaaaaagn acttttttat ngaaaccttt aaattttta 1079

<210> 133
<211> 303
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (295)
<223> a, c, g or t

<400> 133
ttaagtattc aatttctggt ttaaatacca agaggtagaa attaaaggta ggcattggtg 60
tcacagtcca ctaaaaaact agtattccaa cttctattcc ctggcacact actaaatagg 120
caaccaggga tttaaaaaat ggtttctggt gtccaggtaa gtttgcataa aaccaaata 180
aaactgttta atactgggcc cactacatta atctatggtg ctaacacgtg ctgtgaaccg 240
tggggtcagg ggctggggga taaagttgca accatttttt ggggggttgg gggangagga 300
ggg 303

<210> 134
<211> 546
<212> DNA
<213> Homo sapiens

<400> 134
ccggcaaatt taaccacaaa aaaaaagtaa tatgaccata attaatatca gtcaaaatat 60
tcttttaaagg aaaaaaatac taataagaga actctataaa aataaagaat ataataaaaa 120
gagatcacat ttgcaaattt acattgttta atatcatagc ctcaaaataa attgcatata 180
aattttaaaa cctatggaga aattgacaaa tccaccaaca ctgtgggaaa tttttaatac 240
atatctctta gctattaatg cataaagtag gtaaggaaaa ccaataggat gcaaataatt 300
tgaacaataa aatcaacaac tttgatttag ttgatataca tatacagaca cttgcattta 360
gtaattggaa aatatacatt attttccaac acacacaaaa aaacacttgc aaaaatgggc 420
tgtgtcttaa atttttcaaa gaactgatat catacagaac acatgttatg accataatgt 480
agttacatta gaaaatgtgg cagggattct gattctcctt tctgtgctag ggcatacagt 540
taaatac 546

<210> 135
<211> 590
<212> DNA
<213> Homo sapiens

<400> 135
aaaaaagtaa tatgaccata attaatatca gtcaaaatat tcttttaaagg aaaaaaatac 60
taataagaga actctataaa aataaagaat ataataaaaa gagatcacat ttgcaaattt 120
acattgttta atatcatagc ctcaaaataa attgcatata aattttaaaa cctatggaga 180
aattgacaaa tccaccaaca ctgtgggaaa tttttaatac atatctctta gctattaatg 240
cataaagtag gtaaggaaaa ccaataggat gcaaataatt tgaacaataa aatcaacaac 300
tttgatttag ttgatataca tatacagaca cttgcattta gtaattggaa aatatacatt 360
attttccaac acacacaaaa aaacacttgc aaaaatgggc tgtgtcttaa atttttcaaa 420
gaactgatat catacagaac acatgttatg accataatgt agttacatta gaaaatgtgg 480
cagggattct gattctcctt tctgtgctag ggcatacagt taaatcacat tttcaccttc 540
cttgtattta tgagacttag ctctgtcctt atgaatgtgg gcagaagtga 590

<210> 136
<211> 165
<212> DNA
<213> Homo sapiens

<400> 136
gctcgaggcc tggcatctga gttcttctgt tcaggagaaa cactttcagc aggccattga 60
gagggtcac ggaggtgagc ctgggagccc ttagggaggg aggggtgttt gcagctctgg 120
gcttggcagg ctaccccct ggccccagtt tcaattctgc atgca 165

<210> 137
<211> 172
<212> DNA
<213> Homo sapiens

<400> 137
tagttacagt ccttaaataat atgtcttggg tgccctgtgg ctgtgatttt ttaagggaaa 60
ttaacttatt ttaaataaaa taaacttaat ttaaaataaa attttgttat ctaaagccaa 120
atagaaaaaa ttccacattt tttcttacag tgctcattca tcagaacctt tt 172

<210> 138
<211> 809
<212> DNA
<213> Homo sapiens

<400> 138

```
agtacgtaca gtatcaaaca gtctccctcc ttttctctgt gatttgggtct ttctccttag 60
agaatgtcct ccctccaact ccaaaaagaca tgctctctgt gtatagttac agtccttaaa 120
tatatgtcct ggggtgccctg tggctgtgat tttttaaggg aaattaactt attttaaata 180
aaataaaactt aattttaaact aaaattttgt tatctaaagc caaatagaaa aaattccaca 240
ttttttctta cagtgtcat tcatcagaac cttttttttt tcttcttatt ttttcttttt 300
ttggggagaa tgggtcctcc ctttgggtgcg catcaggggg aataagaggt acaaacaggg 360
gggtgattata cgctcacttg ggagtttggg aactccgggg gcatcattgg gattccatt 420
ttgtcctcaa gcctccggag tagctaggac atacgggttt tgcaccacaa ggccgggata 480
aatttcaaaa ttttctcac gagacaaagt ttgggattct tggccccagg attgggacgg 540
gggtatcac aaaagaaact atttcagggg cgcttagaga ggctcaagt acacctactt 600
atcaggggtt tccagtggag agaactgtac cctaccctta ctaccttta agtgggtgct 660
ctccctccac cttaacctt tacacattac ggaactggcg ctatcatttt aaagtcaact 720
aacctggact ttggacttct ttaacacttc agctccggga tccaaactaa aatcttaggc 780
aaggcctaataa ggacggtaga agtctacgc 809
```

<210> 139

<211> 294

<212> DNA

<213> Homo sapiens

<400> 139

```
gtctttttca ttcataagtaa ccctgcaaaa caaacatata gaacagagac attatggaga 60
cttgaggatt gattttatgt attgattatg tatgtaagtc ccgataacat ctctgggtca 120
ggaaattgca agaaaaagat tgggaatcag aacagcagaa aggtattttt ggaagggtaa 180
tttactgatt tttcgtttta aattgttgac attgccttcg ccggtggaaa tgaattactt 240
atgtgaatct ggcaggaaca caatttttaa aattagaaaa ttagtctctc ttat 294
```

<210> 140

<211> 1056

<212> DNA

<213> Homo sapiens

<400> 140

```
acctaaacac attttaatta tattttgtct atttttggag aaccatttcc ctttgacatc 60
tattatgaac attctaaaac ttaaatttgt gaaaacaaaa ctctgggaga tagattgtaa 120
ttttattcca tgaggaaagt gttaaaccag ctttgagtt tgaattttat tcttaaaggc 180
tctgcagttc ttacctggat gtcgaaatga tttttaattt caactgctgt agacctcatc 240
ctgtgggaac tagaaataat gtccaactgc cgtccagtct ggcgacattc cagccgttcc 300
cccacccac gataacggcc tgactcttcc tcaattcatg acagccatt ctacacataa 360
cctttctcct ctggcaccgg tcttcccagc agagagggat cctgcccctc ccttcccact 420
ctccagcata cagaccagca ggaagccaca agagggaaaa acaaaagcct tctgtataag 480
gcctatgaaa ggaccatggg ccagcctcag aatctgctgc ccctacaaac cagtattcct 540
caaatgatag ttccacattt acttaataag gaggactaat tttctaattt taaaaattgt 600
gttctgtcca gattcacata agtaattcat ttccaccggc gaaggcaatg tcaacaattt 660
aaaacgaaaa atcagtaaat tacccttcca aaaatacctt tctgctgttc tgattcccaa 720
tctttttctt gcaatttctt gaaccagaga tgttatcggg acttacatac ataataata 780
cataaaatca atcctcaagt ctccataatg tctctgttct atatgtttgt tttgcagggg 840
tactatgaat gaaaaagaca atttcatgaa tgcagaaaaat ctggggatcg tgtttgggcc 900
cactctgatg aggccccctg aggacagcac cctgaccacc ctgcatgata tgcggtacca 960
aaagctgatt gtgcagattt taatagaaaa cgaagacgtt ttattctaata ccatcagggg 1020
aatgagctga atggccccag caccatccaa gttgac 1056
```

<210> 141

<211> 968

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (319)

<223> a, c, g or t

<220>

<221> unsure

<222> (497)

<223> a, c, g or t

<400> 141

```
acgagatgtc ccagtaacct aaaattatcc agtcggtcct cttactttac aactaagaaa 60
aataaggcctt agaaagaggg attgccagaa actttggcag ctggattgcc tgtgcttggt 120
cctctaagcc atacctaaat tctgcagtaa atacttaact ttttaatagg gaaattgctt 180
caagataact tgaccagtga tacggtaaaa taattagact attggactaa tggtttaaca 240
caagtggcctt taaaaagtct gcttaaaaaa caatttttat ttagaaaaaa tagaaaaata 300
aaaaacatctt caaaatttng gagcctgaag gggctgtttg tttcatatat ggataatctt 360
tgaaaaggca agtcctgtat gtatttttca tttgttgaaa gaagattggg tatcagtagg 420
cttgcaaaaca taatttgctt ttaagttcct tcaaggtttt atgcaataaa acctattgat 480
ttggaacttt aaaaaanaaa acaacaaaaa aatactttca gggtttttga atttcaagt 540
gttttttaag gggagcaata gtttgccatt taccaaaggc ttctccagat aatttcttaa 600
atgtttctac ttaaaaaataa aagctattaa taataagctg tcatgggac catttgaaga 660
cagggaaaat agaaaatttt tattgtaaag ggaagaactt atccttttaa ttttatggac 720
taacagagtc tgcaggtcct aactcatttc agcctgtcaa atgtgcaatt aaaaatgaat 780
tttctaattg tattcaaatg aggcctctata gtgaatacag aatcactcct ctaagttttt 840
tcccagttaa tttgttttaa agtgtgtgac tctcttgcaa gaacgtttaa aagttaagtc 900
ttgtaactgt taacatctaa tgtattaata taagccattt gttttttacc atttttttaa 960
ggccgtat
```

<210> 142

<211> 1466

<212> DNA

<213> Homo sapiens

<400> 142

```
gaaaatttga gtatcttttt gaaattttaa attgaaattt ggatagagat ggttatggag 60
agaaatcaaa caactggaat agctgtttga tatcacttaa aagtgataaa attttaagtt 120
gaatctgggc agtttgcaat ggcctatttg taagaaatat caagacttct tgagaaaaat 180
gaaaagtga tacataaatg cttaaaatct ggtacttctg agttaagggt ttgctctttg 240
agcttaatcc aatttgggat gatttttcat cctagggtt tttgttttcc ttttttattt 300
ttattttttc ttttttttagg ggaaggggac ttgctttcct ttccaaaaag gtgaatcctt 360
cttgtaggac ataggtaaaa aaaacaaagc tgaaatatat gttttgaata tagatagcta 420
attccctggg atataatate ctttcaattt tttttttttt ttgggcccag tctgcctttg 480
gatgtttcaa aagtctgaac gagatgtccc agtaacctaa aattatccag tctgtcttct 540
tactttacaa ctaagaaaaa taaggcttag aaagagggat tgccagaaac tttggcagct 600
ggattgcctg tgcttgttcc tctaagccat acctaaattc tgcagtaaat acttaacttt 660
ttaataggga aattgcttca agataacttg accagtata cggtaaaata attagactat 720
tggaactaat gttaaacaca agtggcctta aaaagtctgc ttaaaaaaca atttttattt 780
agaaaaata gaaaaataaa aacatcttca aaatttagga gcctgaagg gctgtttgtt 840
tcatatatgg ataactttg aaaaggcaag tctgtatgt attttctatt tgttgaaaga 900
agattgggta tcagtaggct tgcaaacata atttgctttt aagttctttc aagggtttat 960
gcaataaaac ctattgattt ggaactttta aaaaaaaaac acaaaaaaaa tactttcagg 1020
gttttgtaat ttcaagtggg tttttaaggg gagcaatagt ttgccattta ccaaaggcct 1080
ctccagataa tttcttaaat gtttctactt aaaaataaaa gctattaata ataagctgtc 1140
atgggatcca tttgaagaca gggaaaatag aaaattttta ttgtaagggt aagaacttat 1200
ccttttaatt ttatggacta acagagtcctg caggtcttaa ctcatttcag cctgtcaaat 1260
```

58

```

gtgcaattaa aatgaatgt tctaattgta ttcaaagag gctctatagt gaatacagaa 1320
tcactcttct aagttttttc ccagttaatt tgtttaaaag tgttgactc tcttgcaaga 1380
acgtttaaaa gttaagtctt gtaactgtta acatctaatt tattaatata agccatttgt 1440
tttttaccat ttttttaagg ccgtat                                     1466

```

<210> 143
 <211> 306
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (289)
 <223> a, c, g or t

```

<400> 143
gacacagcct atctcaaaga gagatgagaa gagccaggcc ccctctcttc ttcctccatg 60
ctgttagctc accagggcag atcttgacct caaagaatgc cgtcttcctt tctggagctg 120
gtcctgtgat gtgaacctgg ctatcttcaa ttcacaggat agggagtaag acatttcatt 180
ttggccttag gtccaagcca tcttcttcaa tgtagctact actagagagc ccacaatgaa 240
gccaataatt ggctccccat ttggcaattt gtgtcccttt cagaaagang aagggttagt 300
aatcac                                             306

```

<210> 144
 <211> 494
 <212> DNA
 <213> Homo sapiens

```

<400> 144
gacacagcct atctcaaaga gagatgagaa gagccaggcc ccctctcttc ttcctccatg 60
ctgttagctc accagggcag atcttgacct caaagaatgc cgtcttcctt tctggagctg 120
gtcctgtgat gtgaacctgg ctatcttcaa ttcacaggat agggagtaag acatttcatt 180
ttggccttag gtccaagcca tcttcttcaa tgtagctact actagagagc ccacaatgaa 240
gccaataatt ggctccccat ttggcaattt gtgtcccttt tcagaaagag gaagggttag 300
taatcagcac ttttaagtac cagcatgcag cattaacaag ttctcaaggc ctgcaagcca 360
tagggtttct gtcttccttg tattggcctt gtaatctctg accatgatta gggttaagagt 420
taagagactc ccaggacagg aaacggaaaa catcagattg tgtatggaat gaacctctct 480
ggctggatgt ggtg                                             494

```

<210> 145
 <211> 174
 <212> DNA
 <213> Homo sapiens

```

<400> 145
gtggaacaac tctatgccat aaaatttctt atttcacagt taaatgaaca tatttgtgtt 60
atgtcacttt ctttttagctt gcattccttt tataggaagg ccatttttagg agtcctgggg 120
cattttgact caacttctta aatcatttat tctattcaca aaaggtttat tgaa      174

```

<210> 146
 <211> 445
 <212> DNA
 <213> Homo sapiens

<220>

<221> unsure
 <222> (371)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (391)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (406)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (427)
 <223> a, c, g or t

<400> 146
 tgatttttaa caattgtgtg tgtgcaccca gctaaccatc tctacaatcg atctagaaca 60
 ttttcatcac ttcagtgtt ctcgtatatt ccttcccagc taacccatga tcccccaacc 120
 tggccatagg aacccgctga tccatcttct atcacttttag attgaatttg tctttcctac 180
 tgttttatat aaagaaatta cctcctttaa gtcctatcaa attcctgac acccttaaaa 240
 aacaattttt aggtattacc ataaaacctt ccatgacatt ctctgcttta tcttctctgt 300
 gctactttgt ccattcattg ttgcattgta atgtatttct gtacatgtta tatcactaaa 360
 ctgtctctc nttgaaggga gggacatgtg ntcacatct atttcaagg cttatacaga 420
 aactganaca tagtagatgc ttact 445

<210> 147
 <211> 734
 <212> DNA
 <213> Homo sapiens

<400> 147
 tgatttttaa caattgtgtg tgtgcaccca gctaaccatc tctacaatcg atctagaaca 60
 ttttcatcac ttcagtgtt ctcgtatatt ccttcccagc taacccatga tcccccaacc 120
 tggccatagg aacccgctga tccatcttct atcacttttag attgaatttg tctttcctac 180
 tgttttatat aaagaaatta cctcctttaa gtcctatcaa attcctgac acccttaaaa 240
 aacaattttt aggtattacc ataaaacctt ccatgacatt ctctgcttta tcttctctgt 300
 gctactttgt ccattcattg ttgcattgta atgtatttct gtacatgtta tatcactaaa 360
 ctgtctctc cttgaaggga gggacatgtg ttcactcatc tattttcaag gcttattaca 420
 gaaactgaaa catagtagat gcttacttgg gaatattata tctcaaaata gaaaaacacc 480
 cagcaaatcg catcttatat tagtcttttag aattagatc aaagcctaatt tattatgaca 540
 cttgaaacat taaataactt agaaaacaaa gacttaaaag ttttatgata aagccagaaa 600
 ctttttatac tgaccatttt taaatactga catttcagat taattggggg cagatgatat 660
 atgaaattat agttttatact gtgacttctt aatacttcag ttgtgttaga taaactgata 720
 gttcgtcaca tttt 734

<210> 148
 <211> 29
 <212> PRT
 <213> Homo sapiens

<400> 148
 Met Leu Lys Ile Ile Asp Lys Leu Tyr Phe Ser Tyr Leu His Ser Ala
 1 5 10 15

60

Asp Ile Leu Cys Asn Thr Glu Ser Tyr Thr Leu Ser Met
 20 25

<210> 149

<211> 87

<212> PRT

<213> Homo sapiens

<400> 149

Met Gly Trp His Glu Ile Gln Ile Pro Val Leu Ile Phe Leu Leu Ala
 1 5 10 15

Val Tyr His Arg Thr Ser His Phe Thr Ser Leu Pro Leu Gly Pro Gln
 20 25 30

Phe Ser Val Phe Leu Ile Tyr Lys Tyr Ser His Pro Ala Phe Arg Gln
 35 40 45

Val Leu Arg Leu Asn Lys Glu Phe Asn Leu Leu Trp Leu His Ile Lys
 50 55 60

His Ile Leu Val Ser Val Cys Leu Val Ile Ser Asn Ala Asn Ile Leu
 65 70 75 80

Ser Ala Pro Cys Pro Glu Cys
 85

<210> 150

<211> 45

<212> PRT

<213> Homo sapiens

<400> 150

Ser Ser Val Ala Leu Ala Leu Gly Ala Leu Thr Val Trp His Ala Val
 1 5 10 15

Leu Ile Ser Arg Gly Glu Thr Ser Ile Glu Arg His Ile Asn Lys Lys
 20 25 30

Glu Arg Arg Arg Leu Gln Ala Lys Gly Arg Val Ser Arg
 35 40 45

<210> 151

<211> 152

<212> PRT

<213> Homo sapiens

<400> 151

Met Val Pro Glu Val Leu Ile Leu Cys His Gly Leu Ala Val Trp Lys
 1 5 10 15

Trp Phe Pro Gly Leu Ala Val Leu Arg Ile Pro Gly Cys Val Thr Gly
 20 25 30

Asn Lys Pro Phe Asn Leu Pro Gly Thr Val Phe Phe Cys Lys Met Arg

61

35 40 45
 Gly Leu Gly Ala Ser Phe Leu Arg Pro Trp Gly Leu Val Ala Glu Phe
 50 55 60
 Ile Ser Pro Thr Pro Cys Pro Ser Ser Tyr Gly Ser Thr His Lys Ala
 65 70 75 80
 Phe His Ser His Lys Glu Lys Ala His Lys Val Pro Gln Pro Pro His
 85 90 95
 Thr Gln Glu Pro His Leu His Pro Ser Leu Lys Ala Arg Leu Pro Leu
 100 105 110
 Pro Gln His Thr Gln Val Leu Leu Gly Leu Pro Ala Leu Phe Ser Ser
 115 120 125
 Ser Pro Glu Trp Asn Gly Pro Ala Met Ala Ser Gln Arg Thr Ala Ser
 130 135 140
 Trp Gln Ser Trp Glu Trp Val Glu
 145 150

<210> 152
 <211> 29
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (14)

<220>
 <221> UNSURE
 <222> (21)

<400> 152
 Met Gly Leu Arg Val Leu Leu Leu Leu Gly Leu Ser Leu Xaa Met Ser
 1 5 10 15
 Gln Lys Pro Leu Xaa Gln Arg Pro Thr Ala Leu Gly Pro
 20 25

<210> 153
 <211> 46
 <212> PRT
 <213> Homo sapiens

<400> 153
 Met Phe Leu Val Glu His Lys Val Cys Ser Gly Asn Thr Gln Val Ser
 1 5 10 15
 Ile Lys Cys Leu Pro Val Val Ser Glu Lys Phe Val Met Lys Tyr Phe
 20 25 30
 Gly Asn Arg Cys Ile Val Ser Val Gly Gly Ala Asp Glu Phe
 35 40 45

62

<210> 154
 <211> 34
 <212> PRT
 <213> Homo sapiens

<400> 154
 Met Thr His Ser Glu Leu Leu Leu Val Ile Thr Ile Asn His Lys Met
 1 5 10 15
 Pro Gln Gly Pro Arg Val Thr Asn Trp Glu Pro Pro Pro Leu Thr Arg
 20 25 30
 Ile Thr

<210> 155
 <211> 99
 <212> PRT
 <213> Homo sapiens

<400> 155
 Met Asp Ser Phe Leu Leu Leu Arg Gln Arg Glu Gly Gly Lys Arg Asn
 1 5 10 15
 Phe Lys Arg Asn Leu Gln Thr Cys Cys Ala Val Gly Pro Thr Gly Ile
 20 25 30
 His Gly Gly Glu Thr Asn Ser Ile Met Leu Leu Gln Ile Leu Leu Lys
 35 40 45
 Lys Gly Phe Asn Cys Leu Thr Lys Tyr Ser Ser Phe Phe His Leu Leu
 50 55 60
 Thr Leu Gln Pro Asn Gln Val Pro His Thr Thr Gly Arg Cys Arg Glu
 65 70 75 80
 Ile Pro Gln Pro Glu Lys Ile Ile His Ala Gly Gln Arg Gln Lys Phe
 85 90 95
 Thr Pro Gly

<210> 156
 <211> 55
 <212> PRT
 <213> Homo sapiens

<400> 156
 Met Gln Phe Leu Leu Cys Leu Ser Leu Leu Asp Phe Phe Ser Ser Thr
 1 5 10 15
 Tyr Lys His Ala Val Met Ser Pro Asn Gln Lys Lys Cys Lys Asn Pro
 20 25 30
 Phe Ser Pro Met Leu Thr His His Pro Ala Val Val Leu Phe Leu Pro

35

40

45

Phe Thr Leu Leu Tyr Tyr Ser
50 55

```
<210> 157
<211> 59
<212> PRT
<213> Homo sapiens
```

<400> 157
Met Leu Gln Val Asp Val Cys Thr Leu Met Val Arg Thr Trp Ser Ser
1 5 10 15

Trp Pro Cys Trp Val Phe Ala Lys Glu Thr Val Leu Cys Ser Trp Gly
20 25 30

Arg Phe His His Leu Ile Arg Ala Val Val Pro Thr Trp Cys Ser Leu
35 40 45

Asp His Leu Tyr Lys Met Phe Ile Gly Gln Gly
50 55

```
<210> 158
<211> 62
<212> PRT
<213> Homo sapiens
```

<220>
<221> UNSURE
<222> (41)

<220>
<221> UNSURE
<222> (57)

<400> 158
Met Thr Lys Arg Met Glu Lys Cys Leu Asn Ile Tyr Lys Arg Leu Asp
1 5 10 15

Val Tyr Arg Gln Ile Val Ser Lys Gly His Arg Ile Val Arg Asn Ser
20 25 30

Val Ile Leu Phe Cys Val Ile Asn Xaa Pro Phe Leu Tyr Pro Phe Thr
35 40 45

Leu Ile Ile Asp Ile His His Phe Xaa Val Ile Ile Gln Leu
50 55 60

```
<210> 159
<211> 47
<212> PRT
<213> Homo sapiens
```

<400> 159
His Leu Asn Arg Phe Ala Asn Ser Val Lys Val Phe Thr Arg Arg His

64

1	5	10	15
Ala Phe Val	Lys Lys Phe Phe Arg	Gly Ser Ala Cys Asn Cys	Ala Glu
	20	25	30
Ser Ala Leu	Leu Ser Ser Gln Leu	Ala His Cys Val	Gly Arg Trp
	35	40	45

<210> 160
 <211> 43
 <212> PRT
 <213> Homo sapiens

<400> 160
Met Gln Glu Ala Glu Gly Arg Leu Asn Lys Pro Gln Gly Gly Arg Val
1 5 10 15
Gly Ala Glu Arg Val Gly Asn Ile Phe Phe Leu Leu Leu Asn Ser Arg
20 25 30
Lys Ala Lys Thr Gln Ser Lys Leu Phe Leu Ser
35 40

<210> 161
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 161
Met Phe Gly Ile Leu Glu Lys Ser Ser Lys Tyr Val His Leu Glu Gly
1 5 10 15
Ser Leu Lys His Pro Val Ile Lys Leu Val Ser Ile Ser Val Val Lys
20 25 30
Asp Glu Tyr Ser Leu Ile Asn Lys Arg Asn Lys Tyr Leu Asn Ser Leu
35 40 45
Thr Ser Ile Leu Asn Arg Phe Cys Gly Gln Met Arg Leu Pro
50 55 60

<210> 162
 <211> 78
 <212> PRT
 <213> Homo sapiens

<400> 162
Met Thr Pro Ala Leu Ala Ala Trp His Val Leu Ile His Pro Asn Val
1 5 10 15
Cys Phe Leu Ala Pro Ala Asp Ser Leu Glu Gly Ser Ile Lys Glu Asp
20 25 30
Trp Val Asn Met Asp Leu Glu Asn Ala His Leu Gln Arg Glu Asn Gly
35 40 45

65

Gly Trp Ala Ala Phe Pro Ser Pro Ala Pro Val Pro Gly Ile Trp Pro
50 55 60

Arg Ser Ala Ser Val Cys Phe Gly Ala Lys Leu Gln Ala Pro
65 70 75

<210> 163
<211> 51
<212> PRT
<213> Homo sapiens

<400> 163
Met Ser Ser Trp Ile Pro Phe Ile Ile Thr Pro Leu Phe Ser Gly Ile
1 5 10 15

Arg Leu Glu Ala Trp Cys Gln Phe Tyr Ser Ser Leu Tyr Pro Phe Ile
20 25 30

His Phe Leu Ser Ile Leu Phe Pro Lys Tyr Phe Phe Ser Ala Pro Ser
35 40 45

Pro Ala Ala
50

<210> 164
<211> 27
<212> PRT
<213> Homo sapiens

<400> 164
Met Gly Ile Ile Pro Lys Cys Met Phe Leu Leu Gln Ser Arg Leu Met
1 5 10 15

Gly Val Ile Thr Asn Thr Ser Leu Leu Leu His
20 25

<210> 165
<211> 52
<212> PRT
<213> Homo sapiens

<400> 165
Met Lys Val Leu Lys Tyr His Asn Glu Ala Cys Gly Phe Tyr Ser Val
1 5 10 15

Val Trp Met Leu Ser Ser Ser Ile Pro Trp Met Pro Thr Gly Met His
20 25 30

Cys Leu Ile Leu Glu Phe Lys Arg Trp Pro Gln Thr Val Arg Leu Ser
35 40 45

Met Trp Pro His
50

<210> 166

66

<211> 47

<212> PRT

<213> Homo sapiens

<400> 166

Met Gly Arg Lys Ser Thr Asn Lys Thr Ala Cys Thr His Ile Asn Thr
 1 5 10 15

Tyr Val Ser Thr Asn Asp Lys Leu Tyr Leu Tyr Arg Ala Trp Glu Gly
 20 25 30

Ser Tyr Ile Thr Leu His Val Ser His Pro Pro His Thr Ser Arg
 35 40 45

<210> 167

<211> 42

<212> PRT

<213> Homo sapiens

<400> 167

Met Cys Trp Gly Tyr Phe Ser Ile Ser Lys Lys Phe Pro Asn Leu Thr
 1 5 10 15

Ser Val Leu Met Asn Leu Gly Thr Asp Leu Ala Val Arg Pro Thr Ser
 20 25 30

Ile Phe Pro Thr Asp Ser Ile Leu Leu Glu
 35 40

<210> 168

<211> 55

<212> PRT

<213> Homo sapiens

<400> 168

Met Asn Lys Ile Lys Gly Lys Ser Val Leu Phe Tyr Met Pro Glu Thr
 1 5 10 15

Ser Arg Ile Phe Arg Lys Val Gln Phe Lys Glu Asn Gln Ala Ala Leu
 20 25 30

Asp Ser Thr Asn Lys Asn Val Ser Leu Ser Glu Glu Leu Val Asn Gln
 35 40 45

Gly Thr Gln Ser Ala Phe Ser
 50 55

<210> 169

<211> 24

<212> PRT

<213> Homo sapiens

<400> 169

Met Met His Met Gln Leu Ile Ser Glu Phe Ser Cys Leu Cys Cys Phe
 1 5 10 15

67

Phe Phe Leu Gly Ile Tyr Ile Lys
20

<210> 170

<211> 68

<212> PRT

<213> Homo sapiens

<400> 170

Met Ile His Leu Ser Glu Val Ser Gly His Leu Lys Glu Arg Lys Gly
1 5 10 15

Lys Ala Ser Cys Gln Lys Gln Lys His Val Leu Tyr Lys Arg Phe Lys
20 25 30

Asn Gln Asn Gly Ile Arg Leu Ser Asn Cys Lys Arg Gln Ser Ser Ala
35 40 45

Phe Lys Ile Leu Arg Lys Asn Asn Val Tyr Ile Lys Ile Phe Ile Ile
50 55 60

Ile Phe Asn Phe
65

<210> 171

<211> 100

<212> PRT

<213> Homo sapiens

<400> 171

Ser Phe Ala Phe Phe Phe Ser Leu Arg Gln Ser Leu Thr Leu Ser Pro
1 5 10 15

Arg Leu Glu Cys Ser Gly Thr Ile Ser Ala His Cys Asn Leu Cys Leu
20 25 30

Leu Gly Ser Ser Asn Ser Ser Ala Ser Ala Ser Gln Val Ala Gly Ile
35 40 45

Thr Gly Thr His His His Ala Gln Val Ile Phe Ile Phe Phe Ile Glu
50 55 60

Met Gly Phe Arg His Ile Gly Gln Ala Gly Leu Lys Leu Leu Thr Ser
65 70 75 80

Gly Asp Pro Pro Ala Ser Ala Ser Glu Ser Ala Gly Ile Thr Gly Val
85 90 95

Arg His His Thr
100

<210> 172

<211> 58

<212> PRT

<213> Homo sapiens

68

<400> 172

Met Glu Cys Leu Ser Ile Asn Leu Thr Lys Asn Val Ser Tyr Leu Tyr
 1 5 10 15

Thr Gly Pro Leu Asn Thr Ser Glu Thr Lys Leu Lys Ser Tyr Leu Ile
 20 25 30

Gly Asn Gln Phe Pro Pro Arg Phe Ile Tyr Arg Val Ser Glu Ile Pro
 35 40 45

Ile Lys Ile Ser Ala Arg Ser Leu Arg Asn
 50 55

<210> 173

<211> 47

<212> PRT

<213> Homo sapiens

<400> 173

Met Asp Lys Glu Glu Ser Ala Val Leu Val Gly Gly Ser Ile Leu Pro
 1 5 10 15

Asp Lys Leu Phe Leu Val Gly Phe Thr Asp Thr Ser Pro Asp Leu Leu
 20 25 30

Pro Ala Ala Thr Val Cys Phe Tyr Asp Ala Cys His His Asp Ile
 35 40 45

<210> 174

<211> 106

<212> PRT

<213> Homo sapiens

<400> 174

Met Thr His Val Gln Leu His Ala Leu Asp Leu Leu Leu Lys Asp Glu
 1 5 10 15

His Lys Ser Glu Ile Ser Thr Pro Trp Gln Pro Tyr Tyr Gln Leu Leu
 20 25 30

Ile Cys Ser Pro His Val Ser Thr Pro Phe Leu Ala Thr Ser Phe Cys
 35 40 45

Pro Ser His Ile Asn Thr Cys Gly Gln Trp Leu Thr Met Leu Lys Leu
 50 55 60

Lys Leu Tyr Pro Asp Glu Ile Leu Lys Arg Asn His Leu Cys Ser Ser
 65 70 75 80

Val Leu Thr Gln Glu Ser Gln His Val Phe Leu Phe Gln Glu Thr Ile
 85 90 95

Ile Ile Cys Thr Asn Ile Tyr Pro Asp Asn
 100 105

<210> 175

69

<211> 35
<212> PRT
<213> Homo sapiens

<400> 175
Met Ser Met Leu Arg Lys Gly Leu Lys Ser Phe Phe Ser Val Cys Val
1 5 10 15
Leu Pro Ser Glu Pro Asn Ile Gly Ile Ser Ala Ser Lys Ile Pro Gln
20 25 30
Gly Gln Glu
35

<210> 176
<211> 54
<212> PRT
<213> Homo sapiens

<400> 176
Met Ser Ser Ser Pro Leu Val Ser Ala Lys Phe Ser Phe Leu Phe His
1 5 10 15
Glu Gly Arg Ala Pro Ser Leu Phe His Pro Leu Met Thr Ser Gln Pro
20 25 30
Leu Glu Phe Cys Leu Met Met Asp Phe Ser Glu Ile Cys Leu Cys Asn
35 40 45
Glu Asp Lys Asp Ser Gly
50

<210> 177
<211> 20
<212> PRT
<213> Homo sapiens

<400> 177
Met Arg Pro Leu Lys Met Ile Arg Thr Ala Lys Lys Leu Phe Val Tyr
1 5 10 15
Leu Gly Ser Tyr
20

<210> 178
<211> 66
<212> PRT
<213> Homo sapiens

<400> 178
Met Met Tyr Tyr Pro Asp Asp Leu Trp Asn Leu Leu Arg Asn Arg Asp
1 5 10 15
Cys Val Ala Phe Leu Ile Met Gly Thr Gly Pro Ser Leu Leu Arg Leu
20 25 30

Pro Met Cys Val Gly Thr Glu Leu Leu Trp His Ser Ser Ser Arg Leu
35 40 45

Met Glu Leu Ser Ser Ser Glu Ala Ser Trp Val Val His Ala Asn Leu
50 55 60

Val Leu
65

```
<210> 179
<211> 70
<212> PRT
<213> Homo sapiens
```

```

<400> 179
Met Cys Val Ile Tyr Gln Arg Gly Ile Cys Asp Glu Lys Lys Asn Leu
  1             5             10             15

Lys Cys Pro Gln Met Phe Gln Leu Ser Glu Thr Glu Lys Thr Leu Thr
      20             25             30

Ser Val Phe Arg Ile Ile Val Ser Asn Ile Leu Lys Ile Asp Val Ser
      35             40             45

Ser Val Met Ile Phe Leu Arg Leu His Gln Arg Thr Ser Leu Asn Leu
      50             55             60

Ser Val Ile Gln Asn Gln
      65             70

```

```
<210> 180
<211> 30
<212> PRT
<213> Homo sapiens
```

```

<400> 180
Met Asn Pro Val Cys Trp Val Gly Phe Gly Glu Val Asn Ile Glu His
  1                      5                      10                     15

Met Glu Phe Lys Tyr Ile Glu Met Asp Thr Val Ile Glu Met
      20                      25                      30

```

```
<210> 181
<211> 55
<212> PRT
<213> Homo sapiens
```

<400> 181
Met His Ala Cys Gly Ser Leu Arg Leu Asp Lys Asp Pro Thr Thr Leu
1 5 10 15
Leu Cys Val Asn Thr Arg Cys Thr Arg Ser His Leu Pro Gly Ala Gly
20 25 30
Gly Trp Trp Arg Lys Val Lys Ser Gln Gln Thr Val His Arg Thr Tyr
35 40 45

Ser Ala Thr Gly Lys Lys Ser
50 55

<210> 182
<211> 16
<212> PRT
<213> Homo sapiens

<400> 182
Met Pro Ala Leu Arg Glu Ala Phe Pro Gln Ala Pro Leu Ala Leu Ala
1 5 10 15

<210> 183
<211> 48
<212> PRT
<213> Homo sapiens

<400> 183
Met Thr Phe Gln Lys Leu Met Ile Leu His Ile His Asp Gln Met Phe
1 5 10 15

Ser Leu Met Glu Ala Ser Asp Val Cys Ser His Gln Ile Arg Phe Lys
20 25 30

Met Ser Val Ser Ser Lys Ser Ser Lys Thr Ser Pro Ser His Gln Lys
35 40 45

<210> 184
<211> 55
<212> PRT
<213> Homo sapiens

<400> 184
Met Ser Val Leu Lys Arg Phe Leu Lys Pro Ser Leu Ser Ile Ala Lys
1 5 10 15

Thr Cys Tyr Val His Tyr Pro Pro Asn Ser Tyr Leu Lys Thr Thr Pro
20 25 30

Lys Met Leu Tyr Phe Val Phe Lys Val Arg Glu Glu Asn Arg Gly Glu
35 40 45

Val Phe Leu Cys Ser Phe Pro
50 55

<210> 185
<211> 14
<212> PRT
<213> Homo sapiens

<400> 185

72

Met Trp Leu Arg Asp Leu Asn Tyr Lys Ile Ala Arg Leu Asp
 1 5 10

<210> 186
 <211> 42
 <212> PRT
 <213> Homo sapiens

<400> 186
 Met Met Phe Phe Tyr Ile Phe Cys Ser Met Gly Leu Leu Ile Pro Phe
 1 5 10 15
 Ser Thr Leu Lys Met Leu Leu Ile Val Phe Pro Leu Ser Leu Phe Pro
 20 25 30
 Lys Arg Asn Leu Leu Ser Phe Leu Ser Leu
 35 40

<210> 187
 <211> 100
 <212> PRT
 <213> Homo sapiens

<400> 187
 Leu Phe Phe Phe Leu Arg Trp Ser Leu Ala Leu Val Thr Gln Ala Gly
 1 5 10 15
 Val Gln Val Val Asp Ile Gly Ser Leu Gln Pro Leu Pro Pro Gly Phe
 20 25 30
 Lys Gln Phe Ser Cys Pro Ser Leu Leu Ser Ser Trp Asp Tyr Arg His
 35 40 45
 Gly Pro Pro Arg Pro Ala Asn Phe Phe Val Phe Leu Val Glu Met Gly
 50 55 60
 Phe His His Val Gly Gln Ala Gly Pro Glu Leu Leu Thr Ser Ser Asp
 65 70 75 80
 Pro Pro Ala Leu Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Ser His
 85 90 95
 Leu Thr Trp Pro
 100

<210> 188
 <211> 106
 <212> PRT
 <213> Homo sapiens

<400> 188
 Met Ser Cys Leu Trp Pro Ser Leu Asp Leu Pro Ser Leu Ser His Ser
 1 5 10 15
 Lys Gln Ser Ser Ser Gln Ala Glu Gly Gln Val Thr Ser His Thr Arg
 20 25 30

73

Gln Arg Phe Pro Asp Gly Ala His Leu His Pro Ser Leu Thr Leu Val
 35 40 45

Leu Ser Gln Asp Ala Pro Leu Arg Leu Ala Pro Pro Thr Leu Cys Leu
 50 55 60

Leu Cys Tyr Trp Ala Ser Leu Pro Ser Pro Arg Thr Pro Glu Leu Leu
 65 70 75 80

Asn Ala Gly Gln Lys Ser Ile Pro Asp Leu Gln Gln Arg His Phe Asp
 85 90 95

Ile Lys Glu Met Ala Leu Asp Phe Cys Leu
 100 105

<210> 189
 <211> 46
 <212> PRT
 <213> Homo sapiens

<400> 189
 Met Val Ile Ser Arg Ile Ser Ile Leu Arg Lys Met Thr Lys Phe His
 1 5 10 15

Lys Phe Cys Ser Gln Leu Thr Glu Pro Gly Arg Arg Thr Gln Pro Lys
 20 25 30

Glu Asn Pro Trp Ser Leu Tyr Asp Thr Asp Trp Leu Glu Lys
 35 40 45

<210> 190
 <211> 46
 <212> PRT
 <213> Homo sapiens

<400> 190
 Met Ser Arg Val Arg Ala Glu Lys Pro Gly Arg Val Ala Lys Leu Ala
 1 5 10 15

Ala Cys Arg Pro Leu Pro Arg Leu Gln Met Ser Gly Ser Ile Pro Leu
 20 25 30

His Lys Cys Lys Glu Lys Ala Ser Met Pro Pro Leu Trp Ser
 35 40 45

<210> 191
 <211> 50
 <212> PRT
 <213> Homo sapiens

<400> 191
 Met Arg Pro Ala Arg Leu Gly Pro Arg Cys Ser Asp Leu Asp Phe Gly
 1 5 10 15

Leu Val Leu Ser Ser Trp Leu Arg Leu Ala Arg Cys Pro Leu Glu Ser

```
<210> 192
<211> 76
<212> PRT
<213> Homo sapiens
```

```
<210> 193
<211> 54
<212> PRT
<213> Homo sapiens
```

```
<210> 194
<211> 27
<212> PRT
<213> Homo sapiens
```

<400> 194
Met Glu Pro Arg Ile Pro Val Lys Thr Phe Ser Phe Asp Lys Arg Ile
1 5 10 15

75

Leu Ile Arg Ile Leu Tyr Gln Ile Glu Gln Lys
 20 25

<210> 195
 <211> 17
 <212> PRT
 <213> Homo sapiens

<400> 195
 Met Leu Gln His Leu Arg Leu Thr Ile Trp Gly Glu Cys Val Trp Val
 1 5 10 15

Phe

<210> 196
 <211> 51
 <212> PRT
 <213> Homo sapiens

<400> 196
 Met Arg Asn Val Ser Leu Ile Ser Cys Glu Asp Ala Asp Phe Thr Glu
 1 5 10 15

Ala Leu Cys Asn Ile Trp Phe Val His Gln Thr Met Leu Ile Asp Cys
 20 25 30

Arg Ser His Leu Leu Pro Arg Trp Leu Thr Lys Thr Val Gly Ser Leu
 35 40 45

Leu Lys Pro
 50

<210> 197
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 197
 Met Ser His Gly Gln Val Leu Gly Asp Val Ala Gly Lys Val Gly His
 1 5 10 15

Ala Leu Gly Thr Glu Asp Gln Thr Phe Ala Val Glu Val Leu Lys Glu
 20 25 30

Thr Thr Pro Phe Phe Arg Ala Ser Ser Gly Pro Thr Gly Asp Pro Trp
 35 40 45

Cys Pro Asp His Lys Ile Gln Ser Lys Pro Val Ser Leu Ser
 50 55 60

<210> 198
 <211> 400
 <212> PRT
 <213> Homo sapiens

<400> 198

```

Met Leu Leu Leu Val Thr Ser Leu Leu Leu Cys Glu Leu Pro His Pro
 1             5             10             15

Ala Phe Leu Leu Ile Pro Glu Lys Ser Asp Leu Arg Thr Val Ala Pro
      20             25             30

Ala Ser Ser Leu Asn Val Arg Phe Asp Ser Arg Thr Met Asn Leu Ser
      35             40             45

Trp Asp Cys Gln Glu Asn Thr Thr Phe Ser Lys Cys Phe Leu Thr Asp
 50             55             60

Lys Lys Asn Arg Val Val Glu Pro Arg Leu Ser Asn Asn Glu Cys Ser
 65             70             75             80

Cys Thr Phe Arg Glu Ile Cys Leu His Glu Gly Val Thr Phe Glu Val
      85             90             95

His Val Asn Thr Ser Gln Arg Gly Phe Gln Gln Lys Leu Leu Tyr Pro
      100            105            110

Asn Ser Gly Arg Glu Gly Thr Ala Ala Gln Asn Phe Ser Cys Phe Ile
      115            120            125

Tyr Asn Ala Asp Leu Met Asn Cys Thr Trp Ala Arg Gly Pro Thr Ala
      130            135            140

Pro Arg Asp Val Gln Tyr Phe Leu Tyr Ile Arg Asn Ser Lys Arg Arg
      145            150            155            160

Arg Glu Ile Arg Cys Pro Tyr Tyr Ile Gln Asp Ser Gly Thr His Val
      165            170            175

Gly Cys His Leu Asp Asn Leu Ser Gly Leu Thr Ser Arg Asn Tyr Phe
      180            185            190

Leu Val Asn Gly Thr Ser Arg Glu Ile Gly Ile Gln Phe Phe Asp Ser
      195            200            205

Leu Leu Asp Thr Lys Lys Ile Glu Arg Phe Asn Pro Pro Ser Asn Val
      210            215            220

Thr Val Arg Cys Asn Thr Thr His Cys Leu Val Arg Trp Lys Gln Pro
      225            230            235            240

Arg Thr Tyr Gln Lys Leu Ser Tyr Leu Asp Phe Gln Tyr Gln Leu Asp
      245            250            255

Val His Arg Lys Asn Thr Gln Pro Gly Thr Glu Asn Leu Leu Ile Asn
      260            265            270

Val Ser Gly Asp Leu Glu Asn Arg Tyr Asn Phe Pro Ser Ser Glu Pro
      275            280            285

Arg Ala Lys His Ser Val Lys Ile Arg Ala Ala Asp Val Arg Ile Leu
      290            295            300

```

77

Asn Trp Ser Ser Trp Ser Glu Ala Ile Glu Phe Gly Ser Asp Asp Gly
 305 310 315 320
 Asn Leu Gly Ser Val Tyr Ile Tyr Val Leu Leu Ile Val Gly Thr Leu
 325 330 335
 Val Cys Gly Ile Val Leu Gly Phe Leu Phe Lys Arg Phe Leu Arg Ile
 340 345 350
 Gln Arg Leu Phe Pro Pro Val Pro Gln Ile Lys Asp Lys Leu Asn Asp
 355 360 365
 Asn His Glu Val Glu Asp Glu Ile Ile Trp Glu Glu Phe Thr Pro Glu
 370 375 380
 Glu Gly Lys Gly Tyr Arg Glu Glu Val Leu Thr Val Lys Glu Ile Thr
 385 390 395 400

<210> 199
 <211> 10
 <212> PRT
 <213> Homo sapiens

<400> 199
 Met Asp Arg Met Glu Lys Arg Gln Thr Asp
 1 5 10

<210> 200
 <211> 20
 <212> PRT
 <213> Homo sapiens

<400> 200
 Met Cys Tyr Ala Thr Leu His Gln Ile Asn Phe Leu Gln Thr Val Leu
 1 5 10 15

Val Pro Gly Leu
 20

<210> 201
 <211> 31
 <212> PRT
 <213> Homo sapiens

<400> 201
 Met Cys Leu Cys Cys Trp Leu Tyr Trp Glu Glu Tyr Gly Pro Leu Ser
 1 5 10 15

Leu Thr Gln Glu Phe His Val Phe Cys Gln Asp Thr Leu His Gly
 20 25 30

<210> 202

78

<211> 54

<212> PRT

<213> Homo sapiens

<400> 202

Met Asn His Ser Leu Ser Ala Phe Gln Arg Ala Leu Gln Val Leu Ile
1 5 10 15

Phe Lys Met Ser Val Tyr Ala Ser Gly Pro Arg Leu Glu Lys Lys Val
20 25 30

Gly Asn Lys Leu Glu Gly Gly Arg Lys Gln Glu Arg Asn Val Thr Tyr
35 40 45

Met Ala Asp Glu Gly Phe
50

<210> 203

<211> 35

<212> PRT

<213> Homo sapiens

<400> 203

Met Ile Lys Ala Tyr His Pro Tyr Phe Glu Asn Phe Asn His Arg Ala
1 5 10 15

Gln Tyr Val Ser Asn Lys Leu Lys Lys Tyr Ser Phe Gln Leu His Phe
20 25 30

Asp Gly His
35

<210> 204

<211> 76

<212> PRT

<213> Homo sapiens

<400> 204

Met Lys Met Val Asn Arg His Met Lys Trp Lys Ser Ser Ala Leu Ser
1 5 10 15

Asp Leu Val Cys Ile Ser Thr Glu Ile Gln Ala Gly Leu Thr Arg His
20 25 30

Thr Ser His Asn Phe Gln Cys His Cys Thr Ile Ile Leu Thr Val Val
35 40 45

Ser Phe Phe Gln Ser Thr Glu Lys Gln Ala Asp Lys Pro Arg His Leu
50 55 60

Asn Val Thr Trp Leu Met Thr Leu Ile Ser Thr Leu
65 70 75

<210> 205

<211> 94

<212> PRT

79

<213> Homo sapiens

<400> 205

Met Glu Gly Gln Asp Ser Leu Arg Asp Val Gly Ala Leu Ser His Leu
 1 5 10 15

Ala His Thr Asp Arg Ser Trp Leu Gly Arg Ala Gly Val Ser Ala Trp
 20 25 30

Arg Pro Ser Ala Ala Gly Asp Pro Gly Phe His Glu Val Gly Gly Val
 35 40 45

His Ala Gly Thr Ser Gln Leu Ala Gly Pro Gly Gly His Pro Gly Gly
 50 55 60

Ala Gly Ala Trp Gly His Glu Phe Thr Lys Val Ala Gln Gly Gln Glu
 65 70 75 80

Glu Thr Val Val Ala Glu Gly Pro Leu Val Glu Ala Trp Ala
 85 90

<210> 206

<211> 53

<212> PRT

<213> Homo sapiens

<400> 206

Met Pro Gln Asp Gln Asp Pro Pro Arg Glu Glu His Ala Ala Leu Arg
 1 5 10 15

Val Leu Phe Pro Arg Val Pro Leu Ala Val Pro His Gln Leu Gly Gly
 20 25 30

Glu Leu Glu Arg Ala Asp Arg Arg Thr Gly Phe Ser Ala Cys Ala Asn
 35 40 45

Ile Leu Thr Cys Pro
 50

<210> 207

<211> 75

<212> PRT

<213> Homo sapiens

<400> 207

Trp Ser Thr Pro Pro Phe Asp Pro Arg Phe Pro Ser Gln Asn Gln Ile
 1 5 10 15

Arg Asn Cys Tyr Gln Asn Phe Leu Asp Tyr His Arg Cys Leu Lys Thr
 20 25 30

Arg Thr Arg Arg Gly Lys Ser Thr Gln Pro Cys Glu Tyr Tyr Ser Cys
 35 40 45

Val Tyr His Ser Leu Cys Pro Ile Ser Trp Val Glu Ser Trp Asn Glu
 50 55 60

80

Gln Ile Lys Asn Gly Ile Phe Ala Gly Lys Ile
 65 70 75

<210> 208
 <211> 44
 <212> PRT
 <213> Homo sapiens

<400> 208
 Met Arg Val Leu Arg Lys Glu Ser Pro Ser Arg His Val Leu Lys Asn
 1 5 10 15

Met Cys Leu Ile Arg Asn Pro Arg Glu Gly Thr Ala Ala Asn Asn Glu
 20 25 30

Met Glu Ser Ala Thr Gly Glu Glu Lys Gly Asn Arg
 35 40

<210> 209
 <211> 83
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (80)
 <223> a, c, g or t

<400> 209
 Met His Arg Lys Lys Lys Leu Glu Ser Phe Leu Leu Leu Ile Pro Pro
 1 5 10 15

Pro Tyr Leu Pro Leu Thr Lys Met Trp Gly Glu Pro Arg Phe Glu Gly
 20 25 30

Ser Thr Gly Pro Cys Pro Gln Asp Ser Met Glu Gln Pro Val Thr Lys
 35 40 45

Gln Gly Ile Ser Leu Lys Ser Cys Leu Pro Lys Lys Ile Lys Leu Pro
 50 55 60

Arg Leu Ala Leu His Pro Ser Pro Pro Arg Ser Phe Pro Leu Lys Xaa
 65 70 75 80

Lys Lys Leu

<210> 210
 <211> 40
 <212> PRT
 <213> Homo sapiens

<400> 210
 Met Thr Arg Phe Ser Gln Ala Ser Ser Ser Lys Asp Lys Thr Pro Pro
 1 5 10 15

81

Leu Pro Ser Met Val Gln Ala Thr Val Leu Val Lys Lys Tyr Ile Phe
 20 25 30

Thr Lys Lys Lys Ser Tyr Val Leu
 35 40

<210> 211

<211> 87

<212> PRT

<213> Homo sapiens

<400> 211

Met Pro Arg Pro Thr Glu Gly Glu Gly Ser Thr Glu Asp Arg Asp Pro
 1 5 10 15

Ile Gly Ile Gln Ser Gln Thr Arg Ala Glu Pro Thr Val Glu Gln Leu
 20 25 30

Met Ser Gly Ala Lys Asp Thr Ser Trp Asn Pro Pro Asp Gly Ser Ser
 35 40 45

Asn Pro Lys Arg Ala Gly Leu Gln Val Gly Leu Asn Trp Arg Asp Pro
 50 55 60

Gln Glu Ser Gly Arg Arg Asn Thr Arg Ala Phe Leu Glu Glu Gly Thr
 65 70 75 80

Phe Ile Leu Asp Ser Asn Gln
 85

<210> 212

<211> 38

<212> PRT

<213> Homo sapiens

<400> 212

Met Met Pro Gly Pro Ala Ala Leu Ile Pro Pro Thr Ala Thr Ala Cys
 1 5 10 15

Leu Leu Val Val Ala Arg Gly Ser Ser Val Pro Lys Asp Ser Ser Leu
 20 25 30

Phe Cys Ile Thr Val His
 35

<210> 213

<211> 88

<212> PRT

<213> Homo sapiens

<400> 213

Met Ser Leu Leu Asp Ala Ser Ser Leu Lys Pro Tyr Asp Ser Phe Leu
 1 5 10 15

Leu Ala Val Leu Phe Leu Thr Arg Asp Asn Lys Gly Phe Ala Ser Gln
 20 25 30

82

Val Cys Met Ala Lys Lys Val Ser Thr Ser Val Asn Gly Ser Phe Leu
 35 40 45

Met Thr Ser Gln Gln Pro Leu Val Lys Asp Val Ile Glu Ile Val Gln
 50 55 60

Arg Leu Gly Ser Val Cys Phe Val Leu Leu Leu Lys Ser Phe His Gly
 65 70 75 80

Ser Lys Leu Phe Leu Ser Ile Val
 85

<210> 214
 <211> 42
 <212> PRT
 <213> Homo sapiens

<400> 214
 Met Val Ile Arg Glu Leu Leu Gly Gly Gln Lys Tyr Pro Asn Pro Val
 1 5 10 15

Gln Gly Arg Asp Pro Trp Thr Val Thr Val Leu Ser Ala Phe Gly Arg
 20 25 30

Glu Gly Asp Ser Glu Ala Gln Thr Arg Ile
 35 40

<210> 215
 <211> 49
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (46)

<400> 215
 Met Pro Asn Cys Ser Val Glu Leu Arg Gly Tyr Tyr Tyr Asn Phe Val
 1 5 10 15

His Tyr Tyr Lys Tyr Phe Ile Leu Val Val Tyr Ser Thr Ala Asp Ser
 20 25 30

Asn Gln Lys Thr Lys Ile Gln Lys Tyr Tyr Ile Leu Glu Xaa Ile Ile
 35 40 45

Met

<210> 216
 <211> 37
 <212> PRT
 <213> Homo sapiens

<220>

83

<221> UNSURE

<222> (6)

<220>

<221> UNSURE

<222> (8)

<400> 216

Met Glu Met Leu Glu Xaa Lys Xaa Thr Ile Ile Asp Ile Val Ser Leu
1 5 10 15

Leu Ala Leu Ser Gly Asp Leu Thr Gln Leu Arg Lys Ser Leu Val Thr
20 25 30

Leu Lys Ile Cys Arg
35

<210> 217

<211> 72

<212> PRT

<213> Homo sapiens

<400> 217

Met Gly Ser Tyr Gly Leu Leu Phe Lys Phe Tyr Gly Ala Ile Phe Thr
1 5 10 15

Ser Val Ala Gln Gly Trp Ser Val Leu His Leu Arg Lys Val Ser Leu
20 25 30

Gly Lys Cys Pro Cys His Pro Ser His Ser Arg Gln Ala Ala Ser Ser
35 40 45

Ala Phe Ser Ser Ser Ser Ser His Ala Trp Ser Ser Pro Phe Val Ile
50 55 60

Phe Ser Ser Leu Thr Pro Ser Leu
65 70

<210> 218

<211> 49

<212> PRT

<213> Homo sapiens

<400> 218

Met Gly Ser Phe Ser Pro Leu Thr Tyr His Leu Gly His Trp Asn Met
1 5 10 15

Ala Ala Cys Gly Ser Val Cys Glu Gly Pro Gly Asp Gly Gln Gly Gly
20 25 30

Ser Ala Leu Phe Cys Phe Tyr Gln His Cys Ser Met Asn Val Phe Leu
35 40 45

Thr

84

<210> 219
<211> 34
<212> PRT
<213> Homo sapiens

<400> 219
Met Leu Thr Arg His His Pro Leu Asn Val Leu Leu His Arg Leu Cys
1 5 10 15

Leu Asn Trp Leu Glu Glu Asn Asn Tyr Pro Arg Asn Thr Asp Tyr Leu
20 25 30

Ile Phe

<210> 220
<211> 34
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (17)

<400> 220
Met Leu Leu Leu Pro Ala Thr Phe Leu Pro Thr Ser His Ala Arg Pro
1 5 10 15

Xaa Gln Pro His Cys His Thr Thr Cys Leu Ile Thr Ser His Val Leu
20 25 30

Thr His

<210> 221
<211> 111
<212> PRT
<213> Homo sapiens

<400> 221
Met Gly Pro Ser Ser Cys Leu Leu Leu Ile Leu Ile Pro Leu Leu Gln
1 5 10 15

Leu Ile Asn Leu Gly Ser Thr Gln Cys Ser Leu Asp Ser Val Met Asp
20 25 30

Lys Lys Ile Lys Asp Val Leu Asn Ser Leu Glu Tyr Ser Pro Ser Pro
35 40 45

Ile Ser Lys Lys Leu Ser Cys Ala Ser Val Lys Ser Gln Gly Arg Pro
50 55 60

Ser Ser Cys Pro Ala Gly Met Ala Val Thr Gly Cys Ala Cys Gly Tyr
65 70 75 80

Gly Cys Gly Ser Trp Asp Val Gln Leu Glu Thr Thr Cys His Cys Gln
85 90 95

85

Cys Ser Val Val Asp Trp Thr Thr Ala Arg Cys Cys His Leu Thr
 100 105 110

<210> 222

<211> 111

<212> PRT

<213> Homo sapiens

<400> 222

Met Gly Pro Ser Ser Cys Leu Leu Leu Ile Leu Ile Pro Leu Leu Gln
 1 5 10 15

Leu Ile Asn Leu Gly Ser Thr Gln Cys Ser Leu Asp Ser Val Met Asp
 20 25 30

Lys Lys Ile Lys Asp Val Leu Asn Ser Leu Glu Tyr Ser Pro Ser Pro
 35 40 45

Ile Ser Lys Lys Leu Ser Cys Ala Ser Val Lys Ser Gln Gly Arg Pro
 50 55 60

Ser Ser Cys Pro Ala Gly Met Ala Val Thr Gly Cys Ala Cys Gly Tyr
 65 70 75 80

Gly Cys Gly Ser Trp Asp Val Gln Leu Glu Thr Thr Cys His Cys Gln
 85 90 95

Cys Ser Val Val Asp Trp Thr Thr Ala Arg Cys Cys His Leu Thr
 100 105 110

<210> 223

<211> 83

<212> PRT

<213> Homo sapiens

<400> 223

Met Asn Val Glu Ala Arg Glu Gln Cys Asp Val Gln Leu Ser Asp Leu
 1 5 10 15

Thr Trp His Leu Ile Trp Leu Glu Val Pro Pro Leu Leu Ser Val Pro
 20 25 30

Trp Leu Trp Ala His Gly Leu Ala Glu Pro Ser Tyr Gly Phe Arg Phe
 35 40 45

Thr Cys Tyr Asn Ile Gln Arg Gln Cys Thr Ser Leu Pro Arg Lys Leu
 50 55 60

Cys Ser Arg His Pro Phe Val Thr Leu Ile Ser Ile Met Asp Thr Thr
 65 70 75 80

Thr Phe Tyr

<210> 224

86

<211> 132
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (3)

<220>
 <221> UNSURE
 <222> (11)

<220>
 <221> UNSURE
 <222> (14)

<400> 224
 Met Asp Xaa Thr Arg Val His Asp Asp Glu Xaa Val Ile Xaa Gly Asp
 1 5 10 15
 Val Phe Val His Glu Val Thr Pro Gly Pro His Arg Trp Val Leu Val
 20 25 30
 Arg Pro Phe Cys Leu Glu Val Arg Ala Val Phe Leu Arg Leu Trp Tyr
 35 40 45
 Tyr Arg Gly Glu Lys Glu Glu Glu Leu Glu Val Arg Glu Arg Ser Cys
 50 55 60
 Arg Leu Gly Arg Cys Asp Gln Gly Gln Arg Asp Gly Val Gln Glu Ala
 65 70 75 80
 Cys Ser Ser Val Ser Cys Ser Leu Arg Gln Glu Val Ser Pro Ser Ser
 85 90 95
 Gln Leu Asp Met Arg Ser Leu Leu Gly Val Pro Leu Ala Glu Val Glu
 100 105 110
 Pro Val Ala Gln His Pro Pro Asn Glu Gly Arg Gly Arg His Leu Gly
 115 120 125
 Gln Cys Leu Leu
 130

<210> 225
 <211> 38
 <212> PRT
 <213> Homo sapiens

<400> 225
 Met Ile Asn Asn Ser Asn His Asn Asn Ser Ser Ser Ser Lys Leu Arg
 1 5 10 15
 Ala Ser Tyr Val Gln Ala Phe Ser Lys His Phe Thr Cys Ile Thr Pro
 20 25 30
 Leu Val Ile Thr Thr Pro
 35

<210> 226
<211> 58
<212> PRT
<213> Homo sapiens

<400> 226
Met Ser Thr Phe Thr Val Leu Lys Asn Thr His Gln Leu Lys Lys Asn
1 5 10 15
Thr Leu Phe Pro Phe Leu Gly His Leu Asn Leu Arg Glu Gln Leu Leu
20 25 30
Tyr Lys Asn Asp Ile Lys Ile Ile His Phe Gly Ser Met Phe Leu Thr
35 40 45
Val Leu Arg Gly Cys Met Val Lys Leu Lys
50 55

<210> 227
<211> 26
<212> PRT
<213> Homo sapiens

<400> 227
Met His Met His Ile Phe Leu Cys Leu Tyr Asn Leu Cys Asn Ile Cys
1 5 10 15
Glu Cys Asn Thr Phe Ser Phe Phe Leu Leu
20 25

<210> 228
<211> 47
<212> PRT
<213> Homo sapiens

<400> 228
Met Leu Asp Val Met Arg Gln Val Ala Arg Ser Trp Leu Thr Ala Met
1 5 10 15
Glu Arg Leu Leu Leu Pro Ala Ala Val Arg Phe Ser Ala Ile Trp Leu
20 25 30
Ala Gly Gln Phe Ala Met Ala Trp Leu Leu Gln Leu Ile Leu Gly
35 40 45

<210> 229
<211> 53
<212> PRT
<213> Homo sapiens

<400> 229
Met Gly Asn Ile Gly Glu Thr Leu Ser Leu Lys Lys Lys Arg Arg Ala
1 5 10 15

88

Gly Gly Glu Ser Val Lys Asp Pro Gly Ser Thr Asp Thr Gly Gly Gln
 20 25 30

Arg Thr Arg Val Gly Val Ser Ser Asn Asp Ser Val Gly Ser Met Gly
 35 40 45

Ala Val Gly Arg Glu
 50

<210> 230
 <211> 80
 <212> PRT
 <213> Homo sapiens

<400> 230
 Met Val Ile Asn Ser Cys Ile Ile Pro Leu Pro Ser Gln Ala Thr Ile
 1 5 10 15

Pro Glu Pro Trp Pro His Gly Ala Cys Ile Phe Arg Ile Gln Thr Pro
 20 25 30

Trp Gly Ser Ser Pro Leu Leu Pro Ser Leu Ser Ser His Pro Leu Thr
 35 40 45

His Leu Ser Cys Tyr Leu Ser Leu Glu Ile Pro Lys Met Met Cys Val
 50 55 60

Met Glu Arg Leu Glu His Gln Leu Gln Asn His Pro Val Thr Leu Ala
 65 70 75 80

<210> 231
 <211> 40
 <212> PRT
 <213> Homo sapiens

<400> 231
 Met Phe Gln Arg Phe Leu Ala Lys Val Thr Val Trp Met Val Val Pro
 1 5 10 15

Leu Thr Lys Thr Ala Met Asn Ala Lys Arg Ala Ser Phe Val Gly Arg
 20 25 30

His Lys Ile Ile Phe Arg Ile Cys
 35 40

<210> 232
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 232
 Met Leu Leu Tyr Leu Ile Thr Arg Gly Asp Val Glu Asn Gly Cys Phe
 1 5 10 15

Ile Phe Ser Val Val Phe Ala Leu
20

<210> 233
<211> 26
<212> PRT
<213> Homo sapiens

<400> 233
Met Pro Pro Arg Gly Leu Pro His Phe Ser Pro His Pro Thr Arg Gln
1 5 10 15

Phe Leu Phe Leu Phe Pro Leu His Thr Lys
20 25

<210> 234
<211> 37
<212> PRT
<213> Homo sapiens

<400> 234
Met Ser Tyr Glu Ile Leu Val Asn Thr Asp Phe Met Ser Pro Phe Leu
1 5 10 15

Arg Thr Leu Leu Val Cys Phe His Leu Tyr Ala Leu Ile Arg Ala Asn
20 25 30

Asn Leu Lys Tyr Pro
35

<210> 235
<211> 40
<212> PRT
<213> Homo sapiens

<400> 235
Met Gly Lys Gly Leu Arg Leu Gly Val Ser Ile Ile Leu Val Lys Ser
1 5 10 15

Phe Phe Thr Tyr Ser Ser Lys Asp Val Asn Tyr Phe Ser Ile His Ser
20 25 30

Asn Ile Lys Ala Val Phe His Phe
35 40

<210> 236
<211> 40
<212> PRT
<213> Homo sapiens

<400> 236
Met Glu Glu Thr Gly Pro Leu Pro Ser Gly Ser Ser Leu Ser Asp Gln
1 5 10 15

90

Gly Glu Thr Ala Leu Ala Leu Gly Asn Ser Arg Ser Asp Gly Gly Arg
 20 25 30

Gln Ser Ser Ser Ser Met Asn Ala
 35 40

<210> 237

<211> 50

<212> PRT

<213> Homo sapiens

<400> 237

Met His Lys Gln Ser Met Ala Arg Ser Ile Leu Arg Ser Pro Leu Gln
 1 5 10 15

Gln Ile Pro Pro Lys Gly Glu Ala Gly Arg Trp Arg Trp Ala Glu Ala
 20 25 30

Ser Cys Val Leu His Thr Phe Ser Thr Ile Leu Asp Phe Leu Phe Phe
 35 40 45

Phe Phe
 50

<210> 238

<211> 49

<212> PRT

<213> Homo sapiens

<400> 238

Ser Ser Trp Gly Asp Ser Phe Ala Val Ser Ala Ala Trp Ala Arg Lys
 1 5 10 15

Gly Ile Glu Glu Trp Ile Gly Arg Gln Arg Cys Pro Gly Gly Val Ser
 20 25 30

Gly Pro Arg Gln Leu Arg Leu Ala Gly Thr Ile Gly Arg Ser Thr Arg
 35 40 45

Glu

<210> 239

<211> 54

<212> PRT

<213> Homo sapiens

<400> 239

Met Leu Arg Pro Leu Thr Val Ala Ser Lys Arg Leu Leu Thr Ile Ser
 1 5 10 15

Thr Leu Lys Ser Pro Leu Val Gly Leu Cys Ser Phe Ser Lys Ser Gly
 20 25 30

Val Leu Arg Glu Gln Ala Leu Phe Ser Ile Ile Asn Leu Ile Asn Thr
 35 40 45

Asp Trp Gln Lys Gln His
50

<210> 240
<211> 23
<212> PRT
<213> Homo sapiens

<400> 240
Met Lys Lys Lys Ser Tyr Pro Asp Lys Ile Asn Gln Cys Phe Ile Phe
1 5 10 15

Leu Glu His Gln Asn Leu Leu
20

<210> 241
<211> 59
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (6) .. (7)

<220>
<221> UNSURE
<222> (9)

<220>
<221> UNSURE
<222> (13)

<220>
<221> UNSURE
<222> (23)

<220>
<221> UNSURE
<222> (27) .. (31)

<220>
<221> UNSURE
<222> (38) .. (40)

<220>
<221> UNSURE
<222> (43)

<220>
<221> UNSURE
<222> (45)

<220>
<221> UNSURE
<222> (47)

92

<400> 241

Met Val Lys Tyr Met Xaa Xaa Leu Xaa Leu Thr Pro Xaa Phe Ser Asn
 1 5 10 15

Leu Leu Gly Thr Leu Lys Xaa Arg Lys Val Xaa Xaa Xaa Xaa Xaa Pro
 20 25 30

Arg Lys Arg Asn Phe Xaa Xaa Xaa Pro Pro Xaa Leu Xaa Lys Xaa Arg
 35 40 45

Cys His Phe Leu His Ile Asp Leu Gln Arg Val
 50 55

<210> 242

<211> 55

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (53)

<400> 242

Met Val Ser Gly Val Gln Val Ser Leu His Lys Thr Lys Ile Lys Leu
 1 5 10 15

Phe Asn Thr Gly Pro Thr Thr Leu Ile Tyr Gly Ala Asn Thr Cys Cys
 20 25 30

Glu Pro Trp Gly Gln Gly Leu Gly Asp Lys Val Ala Thr Ile Phe Trp
 35 40 45

Gly Val Gly Gly Xaa Gly Gly
 50 55

<210> 243

<211> 75

<212> PRT

<213> Homo sapiens

<400> 243

Met Val Ile Thr Cys Val Leu Tyr Asp Ile Ser Ser Leu Lys Asn Leu
 1 5 10 15

Arg His Ser Pro Phe Leu Gln Val Phe Phe Cys Val Cys Trp Lys Ile
 20 25 30

Met Tyr Ile Phe Gln Leu Leu Asn Ala Ser Val Cys Ile Cys Ile Ser
 35 40 45

Thr Lys Ser Lys Leu Leu Ile Leu Leu Phe Lys Leu Phe Ala Ser Tyr
 50 55 60

Trp Phe Ser Leu Pro Thr Leu Cys Ile Asn Ser
 65 70 75

93

<210> 244
<211> 17
<212> PRT
<213> Homo sapiens

<400> 244
Met Ser Trp Val Pro Cys Gly Cys Asp Phe Leu Arg Glu Ile Asn Leu
1 5 10 15

Phe

<210> 245
<211> 30
<212> PRT
<213> Homo sapiens

<400> 245
Met Tyr Val Ser Pro Asp Asn Ile Ser Gly Ser Gly Asn Cys Lys Lys
1 5 10 15

Lys Ile Gly Asn Gln Asn Ser Arg Lys Val Phe Leu Glu Gly
20 25 30

<210> 246
<211> 57
<212> PRT
<213> Homo sapiens

<400> 246
Arg Val Thr Met Asn Glu Lys Asp Asn Phe Met Asn Ala Glu Asn Leu
1 5 10 15

Gly Ile Val Phe Gly Pro Thr Leu Met Arg Pro Pro Glu Asp Ser Thr
20 25 30

Leu Thr Thr Leu His Asp Met Arg Tyr Gln Lys Leu Ile Val Gln Ile
35 40 45

Leu Ile Glu Asn Glu Asp Val Leu Phe
50 55

<210> 247
<211> 70
<212> PRT
<213> Homo sapiens

<220>
<221> UNSURE
<222> (38)

<400> 247
Met Phe Ala Ser Leu Leu Ile Thr Asn Leu Leu Ser Thr Asn Glu Lys
1 5 10 15

Tyr Ile Gln Asp Leu Pro Phe Gln Arg Leu Ser Ile Tyr Glu Thr Asn

94

20

25

30

Ser Pro Phe Arg Leu Xaa Asn Phe Glu Asp Val Phe Ile Phe Leu Phe
 35 40 45

Phe Leu Asn Lys Asn Cys Phe Leu Ser Arg Leu Phe Lys Ala Thr Cys
 50 55 60

Val Lys Pro Leu Val Gln
 65 70

<210> 248

<211> 36

<212> PRT

<213> Homo sapiens

<400> 248

Met Arg Arg Ala Arg Pro Pro Leu Phe Phe Leu His Ala Val Ser Ser
 1 5 10 15

Pro Gly Gln Ile Leu Thr Ser Lys Asn Ala Val Phe Pro Ser Gly Ala
 20 25 30

Gly Pro Val Met
 35

<210> 249

<211> 26

<212> PRT

<213> Homo sapiens

<400> 249

Met Ser Leu Ser Phe Ser Leu His Ser Phe Tyr Arg Lys Ala Ile Leu
 1 5 10 15

Gly Val Leu Gly His Phe Asp Ser Thr Ser
 20 25

<210> 250

<211> 43

<212> PRT

<213> Homo sapiens

<220>

<221> UNSURE

<222> (6)

<400> 250

Met Ser Leu Pro Ser Xaa Arg Arg Gln Phe Ser Asp Ile Thr Cys Thr
 1 5 10 15

Glu Ile His Tyr Asn Ala Thr Met Asn Gly Gln Ser Ser Thr Glu Lys
 20 25 30

Ile Lys Gln Arg Met Ser Trp Lys Val Leu Trp
 35 40

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)